Some contributions to the analysis of skew data on the line and circle

Thesis

UNRESTRICTED

SOME CONTRIBUTIONS TO THE ANALYSIS OF SKEW DATA ON THE LINE AND CIRCLE

by

ARTHUR RICHARD PEWSEY BSc, MSc

Thesis submitted for the degree of Doctor of Philosophy

DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND STATISTICS
THE OPEN UNIVERSITY

OCTOBER 2002

DATE OF SUBMISSION: 9 SEPTEMBER 2002
DATE OF AWARD: 6 DECEMBER 2002

Abstract

In the first part of this thesis we consider the skew-normal class of distributions on the line and its limiting general half-normal distribution. Inferential procedures based on the methods of moments and maximum likelihood are developed and their performance assessed using simulation. Data on the strength of glass fibre and the body fat of elite athletes are used to illustrate some of the inferential issues raised.

The second part of the thesis is devoted to a consideration of the analysis of skew circular data. First, we derive the large-sample distribution of certain key circular statistics and show how this result provides a basis for inference for the corresponding population measures.

Next, tests for circular reflective symmetry about an unknown central direction are investigated. A large-sample test and computer intensive variants of it are developed, and their operating characteristics explored both theoretically and empirically. Subsequently, we consider tests for circular reflective symmetry about a known or specified median axis. Two new procedures are developed for testing for symmetry about a known median axis against skew alternatives, and their operating characteristics compared in a simulation experiment with those of the circular analogues of three linear tests. On the basis of the results obtained from the latter, a simple testing strategy is identified. The performance of the tests against rotation alternatives is also investigated. Throughout, the use of the various tests of symmetry is illustrated using a wide range of circular data sets.

Finally, we propose the wrapped skew-normal distribution on the circle as a potential model for circular data. The distribution's fundamental properties are presented and inference based on the methods of moments and maximum likelihood is explored. Tests for limiting cases of the class are proposed, and a potential use of the distribution is illustrated in the mixture based modelling of data on bird migration.

This thesis is dedicated with great respect, gratitude and affection

to

Professor Toby Lewis

Acknowledgements

As I have grown older, I have become increasingly more reflective upon my past. The physical distance presently separating me from my roots only heightens the competing senses of melancholy, for what has been, and deep gratitude to those from whom I have learnt. Sadly, it is too late to express in person my gratitude to some who are most deserving of it. Nevertheless, I would like to take this opportunity to express the sincere gratitude I feel to those who have directly contributed towards my knowledge and understanding of Mathematics and Statistics, of which this thesis is an expression.

Going back in time at least 25 years, two people who influenced my mathematical development greatly were my father and my 'A' level mathematics teacher Mr Derek Friggins. My father skillfully used evening washing up sessions to instill in me basic concepts of algebra. On deriving the theoretical results of Chapter 3, I was overcome by sensations similar to those I had felt on obtaining beautifully simple solutions to some initially horrendous problems in Applied Mathematics involving interminable trigonometric expressions. The late Mr Friggins was the person who set those problems and taught me the mathematics required to solve them.

Moving on to my years as a university student in Hull, I am infinitely grateful to those who extended my mathematical knowledge and introduced me to the world of Statistics. I received my first course in Statistics from Edward Evans, and he and Michael Bingham kindled my initial interest in things statistical. In my second and third years, the stimulating lectures and tutorials given by Michael Bingham, Carolyn Craggs, Fergus Daly, Michael Goldstein, Wilfred Kendall and Toby Lewis heightened that interest still further.

For the influence they had on my wider understanding of Statistics whilst a research student in Newcastle, I would like to express my gratitude to the late Jon Anderson, Robin Plackett and, especially, Peter Diggle.

I would also like to thank my former colleagues Michael Bingham, Michael Goldstein and Jim Thompson, in Hull, and Ray Harris, Martin Roberts, Tony Scallan and Chris Sutton, in Preston, for their multifarious contributions to my greater knowledge and understanding of Statistics.

However, this thesis is also a testament to the help and support of many other people. I am greatly indebted to Kevin McConway who, whilst Head of the Department of Statistics, saw through my application to study as a research

student. I thank both him and his successor, Paul Garthwaite, for co-ordinating the administrative and managerial requirements of my studies. Regarding the other forms of support I have received during the research which led to the production of this thesis, I would like to express my sincere gratitude to:

i) The Open University for its excellent distance learning support and on-line library facilities.

ii) Anunciación Gutiérrez, Encarna Pegado, Chris Sutton and my supervisors Chris Jones and Toby Lewis for tracking down sources in the literature.

iii) Tony Scallan and David Wooff for pointers to software archives.

iv) Bruno Bruderer of the Swiss Ornithological Station in Sempach, Switzerland, for providing me with the data analyzed in Chapter 6.

v) Adelchi Azzalini for letting me have access to a pre-publication version of Azzalini & Capitanio (1999) and for stimulating communication by e-mail.

vi) Leon Harter for his generous guidance to published work related to the general half-normal distribution.

I have all sorts of reasons to be grateful to my wife, Lucía Aguilar Zuil. It was she who enticed me to Spain, a move which acted as a catalyst to renew my interest in research. She has been my emotional pillar of strength throughout.

There are three people whose influence on the production of this thesis has been absolutely fundamental: Nick Fisher, Chris Jones and Toby Lewis. The original idea of wrapping the skew-normal distribution on to the circle so as to produce a model for skew circular data is due to Nick Fisher who posed it as a problem to my supervisor Toby Lewis to be cogitated over accompanied by an early morning brandy. All the other ideas expressed within the six chapters making up this thesis can, in fact, be traced back to that one original idea. So, thank you, Nick!

The supervision I have received from Chris Jones has been truly excellent. Mostly our communication has been via e-mail and post, and from the experience of using these forms of communication I have been convinced of the efficacies of both and the enormous potential of conducting research "at a distance". Rarely did I have to wait more than half an hour for a response to any e-mail sent by me. Usually within a week of posting the draft of a paper or a chapter of this thesis, I would receive an e-mail with concise and insightful feedback. The contents of my published papers and this thesis have benefitted enormously from that feedback. Thank you for all your hard work, Chris!

I find it incredibly difficult to know where to start in expressing my gratitude to Toby Lewis. I have known Toby for nearly 25 years and during that time he has been an inspirational lecturer, friend, potential supervisor, role model, confidant, supervisor and guru to me. Always positive, always supportive, always young at heart. I feel immensely privileged that in 1978 our paths in life crossed. For all these reasons, and more, Toby, this thesis is dedicated to you. Thank you so very much.

Publications Arising from this Thesis

Pewsey, A. (2000) Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, **27**, 859-870.

Pewsey, A. (2000) The wrapped skew-normal distribution on the circle. *Communications in Statistics – Theory & Methods*, **29**, 2459-2472.

Pewsey, A. & *Aguilar Zuil, L. (2000) Estimación para la distribución skew-normal de Azzalini con parametrización centrada. In the proceedings of the *XXV Congreso Nacional de Estadística e Investigación Operativa, Vigo, Spain*, pp. 145-146.

Pewsey, A. & *Aguilar Zuil, L. (2000) Contrastes para detectar diferencias con la distribución half-normal dentro de la clase skew-normal. In the proceedings of the *XXV Congreso Nacional de Estadística e Investigación Operativa, Vigo, Spain*, pp. 507-508.

Pewsey, A. & *Aguilar Zuil, L. (2001) The wrapped skew-normal distribution: a model for circular data. In the proceedings of the *XXVI Congreso Nacional de Estadística e Investigación Operativa, Úbeda, Spain*, p. 16.

Pewsey, A. & *Aguilar Zuil, L. (2001) An application of the wrapped skew-normal distribution in the mixture based modelling of bird flight data. In the proceedings of the *XXVI Congreso Nacional de Estadística e Investigación Operativa, Úbeda, Spain*, p. 151.

Pewsey, A. (2002) Large-sample inference for the general half-normal distribution. *Communications in Statistics – Theory & Methods*, **31**, 1045-1054.

Pewsey, A. (2002) Testing circular symmetry. *The Canadian Journal of Statistics*. Accepted for publication.

* The name of my wife, Lucía Aguilar Zuil, appears in these references in recognition of her contribution to the translation into Spanish of the material presented at the two conferences in question. In all other respects, the intellectual ownership of these publications is mine.

Contents

Page

Page

Page

Glossary of Notation

Relations

$\approx$	is approximately equal to
$\sim$	is distributed as

Population quantities

Part I

μ	mean
σ	standard deviation
γ_1	coefficient of skewness
γ_2	coefficient of kurtosis
μ_k	k^{th} central moment about the mean
β_2	μ_4/μ_2^2
β_3	$\mu_5/\mu_2^{5/2}$
β_4	μ_6/μ_2^3

Part II

μ	mean direction
$\tilde{\mu}$	median direction
ρ	mean resultant length
α_p, β_p	p^{th} trigonometric moments about the origin
$\overline{\alpha}_p, \overline{\beta}_p$	p^{th} trigonometric moments about the mean direction
α_p^*, β_p^*	p^{th} trigonometric moments about the median direction
s	circular measure of skewness
k	circular measure of kurtosis
v	circular variance

Sample quantities

Part I

$\overline{y}$	mean
s	standard deviation
g_1	coefficient of skewness
m_k'	k^{th} moment about the origin
m_k	k^{th} central moment about the mean
$\tilde{\mu}$	method of moments estimate of μ
$\hat{\mu}$	maximum likelihood estimate of μ

Part II

$\overline{\theta}$	mean direction
$\tilde{\theta}$	median direction
$\overline{\theta}^*$	$\overline{\theta} - \tilde{\mu}$
$\overline{R}$	mean resultant length
a_p, b_p	p^{th} trigonometric moments about the origin
$\overline{a}_p, \overline{b}_p$	p^{th} trigonometric moments about the mean direction $\overline{\theta}$
a_p^*, b_p^*	p^{th} trigonometric moments about the median direction $\tilde{\mu}$
$\tilde{s}$	circular measure of skewness
$\tilde{k}$	circular measure of kurtosis

Distributions

$\text{SN}(\lambda)$	standard skew-normal with skewness parameter λ
$\text{SN}_{\text{D}}(\xi,\eta,\lambda)$	general skew-normal with direct parameters ξ, η and λ
$\text{SN}_{\text{C}}(\mu,\sigma,\gamma_1)$	general skew-normal with centred parameters μ, σ and γ_1
$\text{SNE}_{\text{D}}(\xi,\eta,\lambda,\zeta)$	extended general skew-normal with direct parameters ξ, η, λ and ζ
$\text{HN}(\xi,\eta)$	half-normal with parameters ξ and η
$\text{WSNC}_{\text{D}}(\xi,\eta,\lambda)$	wrapped skew-normal on the circle with direct parameters ξ, η and λ
$\text{WSNC}_{\text{C}}(\mu_L,\sigma,\gamma_1)$	wrapped skew-normal on the circle with centred parameters μ_L, σ and γ_1

Functions

$\tan^{-1}$ (or arctan)	Consider the two reals x and y, and the ratio $r = y/x$. If $x = y = 0$ then r and $\tan^{-1}(y/x)$ are undefined. Otherwise, $$\tan^{-1}(y/x) = \begin{cases} \theta & \text{if } x \geq 0 \\ \theta + \pi & \text{if } x < 0 \end{cases},$$ where $\theta = \tan^{-1}(r)$ takes values in $[-\pi/2, \pi/2]$.
$\Im$	defined on page 164

List of Figures and Tables

Page

Figures

Page

FIGURES (continued)

Page

FIGURES (continued)

TABLES

Page

TABLES (continued)

Introduction

1 Overview

The content of this thesis has been divided into two parts; one of which considers inference for the general skew-normal class of distributions and its limiting general half-normal distribution, the other consisting of contributions to the analysis of skew data on the circle. However, whilst the types of data to which the two parts refer are different, it is important to stress that there are certain key ideas common to both. One such key concept is that of symmetry, or, perhaps more correctly, a lack of it. For many, symmetry is an aesthetical necessity. Indeed, this role of symmetry manifests itself strongly in the surviving cultural heritage dating back at least to the times of the ancient Greeks. From a mathematical perspective, an assumption of symmetry generally results in greater tractability. Of course, the best known and most frequently applied model in Statistics, the normal distribution, is a symmetric one. Whilst the theoretical importance of the normal distribution is beyond question, the following observation quoted from Pearson (1900) raises considerable doubt as to the role of the normal distribution as a model for real linear data.

> *"We can only conclude from the investigations here considered that the normal curve possesses no special fitness for describing errors or deviations such as arise either in observing practice or in nature."*

For the analysis of circular data, the best known and most frequently applied model is the von Mises distribution, again, a symmetrical one. However, the following remark quoted from Mardia (1972, p. 10) raises questions as to what the true role of the von Mises distribution should be.

> *"As on the line, symmetrical distributions on the circle are comparatively rare."*

In the analysis of skew linear data, a frequently applied technique is that of Box-Cox transformation. For data considered initially to be non-normal, a transformation is sought which, amongst producing other desirable outcomes, results in a transformed data set for which the normal distribution is a reasonable model. Inference is then performed for the transformed data, with results for the original scale usually being obtained by applying the appropriate inverse transformation. Despite the potential problems associated with the application of such an approach being well documented (see, for example, Chatfield (1995, p. 69) and Aitkin *et al.* (1989, Section 3.1)), the (mis)use of this technique is commonplace. Of course, those freed from the shackles of normality have, at least in theory, a vast array of alternative models for linear data at their disposal. The skew-normal class and its limiting general half-normal distribution considered in Part I of this thesis form part of that bank of models.

For circular data distributed over anything other than a reduced arc of the unit circle, the application of any transformation of the data which changes the relative positions of the data can result in wildly different inferential results. Due to the compactness on the circle, the only types of transformation which leave the relative positions of the data unaffected are those of rotation and reflection. Consequently, the only viable general approach to modelling circular data is to fit appropriate models to the original data values or those obtained after applying a rotation or reflection to them. And if, as has been commented, symmetrically distributed circular data are rare, then models capable of describing the forms of asymmetry manifested by real circular data are required. How strange, then, to find that, in the sum total of over 1400 pages making up the five principal texts which address the analysis of circular data, i.e. Mardia (1972), Batschelet (1981), Fisher (1993), Mardia & Jupp (1999) and Jammalamadaka & Sengupta (2001), less than ten are specifically devoted to asymmetric models. In an initial attempt to address this disjuncture between established practice in the statistical analysis of circular data and the implicit statistical demands raised by real circular data, in Chapter 6 we propose the wrapped skew-normal distribution on the circle as a potential model for circular data.

As is also evident from a consideration of the literature, methods are required which can be used in the analysis of circular data to detect symmetry or the lack of it. In Chapters 4 and 5 we propose procedures for testing for two types of symmetry associated with circular data. The remaining chapter of Part II, Chapter 3, provides the theoretical results underpinning much of Chapters 4, 5 and 6.

2 Computing

In order to implement the methodology developed within this thesis, and to explore the sampling properties of estimators, the operating characteristics of tests and the coverage of confidence sets, etc., it was necessary to develop a substantial library of computer programs. Those programs are available from the author upon request. All programming was conducted in FORTRAN following the advice regarding style and readability promoted by Ellis (1990).

Many of those programs make use of two highly efficient sorting routines; one written for sorting a single array by James J. Filliben, the other for sorting two arrays written by Rondall E. Jones and John A. Wisniewski. Both of these routines are freely available over the internet from the GAMS (Guide to Available Mathematical Software) archives.

The pseudo-random number generator routine used was a double precision version of that proposed by Wichmann & Hill (1982), incorporating the amendment of McLeod (1985). The numerical optimization of log-likelihood functions was performed using a routine for the Nelder-Mead simplex (Nelder & Mead, 1965) coded by O'Neill (1971). Values of the standard normal distribution function were evaluated using a routine written by Alan J. Miller. All three routines are available over the internet from the StatLib archives.

Finally, all graphical work was produced using the facilities of the Minitab statistical software package.

3 Notation

Generally, the notation employed in Part I of this thesis follows established convention. That used to denote the different parametrizations of the skew-normal distribution differs from the notation commonly used in the associated literature. We believe the notation used here is less prone to misinterpretation.

In Part II, we follow, as far as is possible, the notation established in Mardia & Jupp (1999). The specific notation used for the three parametrizations of the wrapped skew-normal distribution on the circle is an extension of that used by us for its linear analogue.

PART I

INFERENCE FOR THE SKEW-NORMAL CLASS OF DISTRIBUTIONS AND ITS LIMITING GENERAL HALF-NORMAL DISTRIBUTION

Chapter 1 Problems of Inference for the Skew-normal Distribution

1.1 Introduction

In this chapter we consider the general skew-normal distribution and certain problems of inference associated with it. The distribution's genesis is reviewed in Section 1.2, and in Section 1.3 we briefly discuss the extended general skew-normal distribution.

The main original work presented in the chapter appears in Section 1.4, in which we address problems of inference for the so-called direct and centred parametrizations of the general skew-normal distribution. We discuss these problems from the perspectives of moment and likelihood based inference. Certain irregular features of the sampling distributions of the moment estimators, and others of the log-likelihood surface, are traced by us to the parameter redundancy of the direct parametrization.

In Section 1.4.2 we illustrate the improvements brought about by the centred parametrization. What we believe to be a new general asymptotic result for the joint sampling distribution of the sample mean, standard deviation and coefficient of skewness is given in Theorem 1.1 of Section 1.4.2.1.

In Section 1.4.2.2 we propose a numerical approach to finding the maximum likelihood estimates employing a constrained version of the log-likelihood and a grid based search incorporating the Nelder-Mead simplex. We also detail shortcomings of two S-PLUS routines for likelihood based estimation developed by Azzalini & Capitanio (1999).

The results from a simulation study designed to explore and compare the small-sample characteristics of the method of moments and maximum likelihood estimators for the centred parameters are given in Section 1.4.2.3.

Tests for limiting cases of the distribution are discussed in Section 1.4.3. We propose two new procedures for testing the null hypothesis of an underlying half-normal distribution against the alternative of some less positively skewed member of the skew-normal class. The first is based on an asymptotic result for the

sampling distribution of the coefficient of skewness, whilst the second is a Monte Carlo variant of it founded upon the same statistic.

In Section 1.4.4 we illustrate the developed methodology in the analysis of two real data sets.

For the most part, the content of Section 1.4 follows closely that of Pewsey (2000a). However, the level of detail given within it is generally greater than that in the cited paper. Also, Theorem 1.1 is a generalization of a result given in the same paper for the sampling distribution of the moment estimators of the centred parameters.

The chapter closes with a summary of its content and a description of potential avenues for related future research.

1.2 Definition, Genesis and Properties of the Skew-normal Class

1.2.1 The Standard Skew-normal Distribution

The (standard) skew-normal distribution, popularly referred to as "Azzalini's skew-normal distribution", see Johnson *et al.* (1994, p. 61), was developed formally by Azzalini (1985) as a particular case arising from a variant of the following lemma. Rather than repeat the rather opaque proof of the lemma given by Azzalini (1985), we present what we consider to be a more transparent alternative.

Lemma 1.1 *Consider two arbitrary absolutely continuous distributions with densities f and g and distribution functions F and G, respectively, which are symmetric about 0. Then, for any $\lambda \in (-\infty, \infty)$,*

$$2\,G(\lambda y)\,f(y), \qquad (-\infty < y < \infty) \tag{1.2.1}$$

is a probability density function.

Proof Given the assumed symmetry about 0, it follows that

$$\begin{aligned}\int_{-\infty}^{\infty} 2G(\lambda y)f(y)dy &= 2\left\{\int_{-\infty}^{0} G(\lambda y)f(y)dy + \int_{0}^{\infty} G(\lambda y)f(y)dy\right\} \\ &= 2\left[\int_{-\infty}^{0} G(\lambda y)f(y)dy + \int_{-\infty}^{0}\{1-G(\lambda y)\}f(y)dy\right] \\ &= 2\int_{-\infty}^{0} f(y)dy = 1.\end{aligned}$$

With the aim of defining a class of distributions which includes the standard normal, is mathematically tractable and includes distributions with wide-ranging coefficients of skewness and kurtosis, Azzalini (1985) considered the model arising from (1.2.1) with $f(\cdot)=\phi(\cdot)$ and $G(\cdot)=\Phi(\cdot)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ represent the probability density function (pdf) and distribution function, respectively, of the standard normal distribution. These choices lead to the following definition of a standard skew-normal random variable.

If a random variable X has pdf

$$f(x;\lambda)=2\phi(x)\Phi(\lambda x)\,, \qquad -\infty<x<\infty, -\infty<\lambda<\infty, \tag{1.2.2}$$

then X is said to be distributed according to the standard skew-normal distribution with parameter λ, denoted $X\sim \mathrm{SN}(\lambda)$.

The parameter λ regulates the skewness of the distribution. Azzalini (1985) showed that the distribution function of X can be represented as

$$F(x;\lambda)=\Phi(x)-2T(x,\lambda),$$

where $T(h,a)$, a function studied by Owen (1956), gives the integral of the bivariate standard normal density over the region bounded by the lines $x=h$ and $y=ax$ in the (x, y) plane.

Henze (1986) gave a probabilistic representation of the standard skew-normal distribution which reveals the structure of the skew-normal class. He showed that if Z_1 and Z_2 are independent standard normal random variables then

$$X=\frac{\lambda}{(1+\lambda^2)^{1/2}}|Z_1|+\frac{1}{(1+\lambda^2)^{1/2}}Z_2=\delta|Z_1|+(1-\delta^2)^{1/2}Z_2\sim \mathrm{SN}(\lambda), \tag{1.2.3}$$

where $\delta=\lambda/(1+\lambda^2)^{1/2}\in(-1,1)$. Thus, a SN($\lambda$) random variable can be viewed as a normalized linear combination of independent half-normal and standard normal random variables. Corollary 2 of Henze (1986), which is based on (1.2.3) and an adaptation of the Box-Müller method, provides an efficient means of generating random variables from the SN(λ) distribution. This representation serves as a means of relating the skew-normal distribution to the distributions of other sums of random variables studied prior to the publication of Azzalini (1985). The distribution of the sum of a normal random variable and a truncated normal

random variable was considered in Weinstein (1964) and Nelson (1964). In the econometrics literature, Aigner *et al.* (1977) derived the distribution of the sum of a normal random variable and a half-normal random variable.

Two other constructions which lead to the standard skew-normal distribution are also worthy of note. Both are associated with the following bivariate set-up. Consider the joint distribution of (Z_1, Z_2) where Z_1 and Z_2 are two standard normal random variables with $\text{corr}(Z_1, Z_2) = \rho$. Let $Z_{(1)}$ denote the minimum of Z_1 and Z_2, and $Z_{(2)}$ the maximum.

The first construction predates the work of Azzalini (1985) and follows from a more general result quoted by David (1981, pp. 117-118) as being a private communication due to Nagaraja (1979). For a_1, $a_2 = -\rho a_1 \pm \{a_1^2(\rho^2 - 1) + 1\}^{1/2}$ and $a_1 + a_2$ all non-zero, the linear combination $Y = a_1 Z_{(1)} + a_2 Z_{(2)}$ is a standard skew-normal random variable with $\lambda = \pm\{(1-\rho)/(1+\rho)\}^{1/2}\{(a_2 - a_1)/(a_1 + a_2)\}$, the sign depending on the sign of $a_1^{-1} + a_2^{-1}$. In passing, we note that Loperfido (2002) considered the distributions of $Z_{(1)}$ and $Z_{(2)}$ in this context, showing them to be $\text{SN}(-\lambda^*)$ and $\text{SN}(\lambda^*)$, respectively, where $\lambda^* = \{(1-\rho)/(1+\rho)\}^{1/2}$. In fact, the result for the distribution of $Z_{(1)}$ had been derived much earlier by Roberts (1966).

The second construction follows from a more general one given by Arnold *et al.* (1993). If Z_2 is truncated below at 0 then the marginal distribution of Z_1 is standard skew-normal with $\lambda = \rho/(1-\rho^2)^{1/2}$.

We also note that Andel *et al.* (1984) identified the standard skew-normal distribution as the stationary distribution of a certain first-order threshold autoregressive process: see, also, Azzalini (1986).

1.2.2 Fundamental Properties of the Standard Skew-normal Distribution

The basic properties of the standard skew-normal distribution were investigated by Azzalini (1985, 1986) and Henze (1986). The most important five can be summarized as follows:

a) When $\lambda = 0$, the distribution corresponds to the standard normal distribution.

b) As $\lambda \to \infty$, the distribution tends to the positive standard half-normal distribution. Conversely, as $\lambda \to -\infty$, the distribution tends to the negative standard half-normal distribution.

c) If X is distributed according to the standard skew-normal distribution with parameter λ, i.e. $X \sim \mathrm{SN}(\lambda)$, then $-X \sim \mathrm{SN}(-\lambda)$.

d) The density (1.2.2) is strongly unimodal.

e) If $X \sim \mathrm{SN}(\lambda)$, then $X^2 \sim \chi^2_1$.

Densities for various choices of positive λ, ranging between N(0,1) ≡ SN(0) and the limiting standard half-normal ≡ SN(∞), are given in Figure 1.1. From this plot we see that, for $\lambda = 2$ the density is moderately asymmetric, whilst for $\lambda = 20$ the density differs little from that of the limiting standard half-normal distribution.

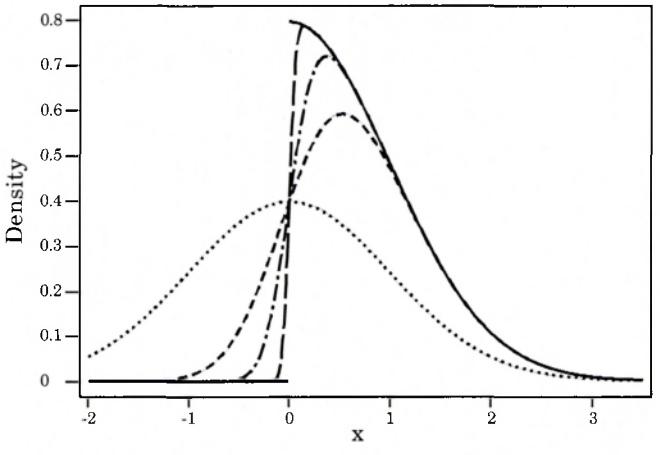

Figure 1.1 Densities for various cases of the standard skew-normal distribution: · · · · · , SN(0) ≡ N(0,1); – – – –, SN(2); — · — , SN(5); —— ——, SN(20); ————, SN(∞) ≡ standard half-normal.

Azzalini (1985) obtained the moment generating function of $X \sim \mathrm{SN}(\lambda)$ as $M(t) = 2\exp\left(t^2/2\right)\Phi\left(\delta t\right)$. Given property e) above, the even moments of X are equal to the even moments of a standard normal random variable. Henze (1986) used the representation (1.2.3) to obtain the following expression for the odd moments of X.

$$E\left(X^{2k+1}\right) = b\lambda\left(1+\lambda^2\right)^{-(k+1/2)} 2^{-k}\left(2k+1\right)! \sum_{\nu=0}^{k} \frac{\nu!\left(2\lambda\right)^{2\nu}}{\left(2\nu+1\right)!\left(k-\nu\right)!}, \qquad (1.2.4)$$

where $b = (2/\pi)^{1/2}$.

1.2.3 The General Skew-normal Class

The obvious extension of the standard skew-normal class results from the inclusion of location and scale parameters. Thus, if $X \sim \text{SN}(\lambda)$ then $Y_{\text{D}} = \xi + \eta X$ $(-\infty < \xi < \infty; \eta > 0)$ is said to have a (general) skew-normal distribution with what Azzalini & Capitanio (1999) refer to as "direct" parameters (ξ, η, λ). We denote the fact as $Y_{\text{D}} \sim \text{SN}_{\text{D}}(\xi, \eta, \lambda)$, the subindex D referring to the use of the direct parametrization. From (1.2.2), the density of Y_{D} is given by

$$f(y;\xi,\eta,\lambda) = \frac{2}{\eta}\phi\left(\frac{y-\xi}{\eta}\right)\Phi\left\{\lambda\left(\frac{y-\xi}{\eta}\right)\right\} \quad (-\infty < y < \infty). \tag{1.2.5}$$

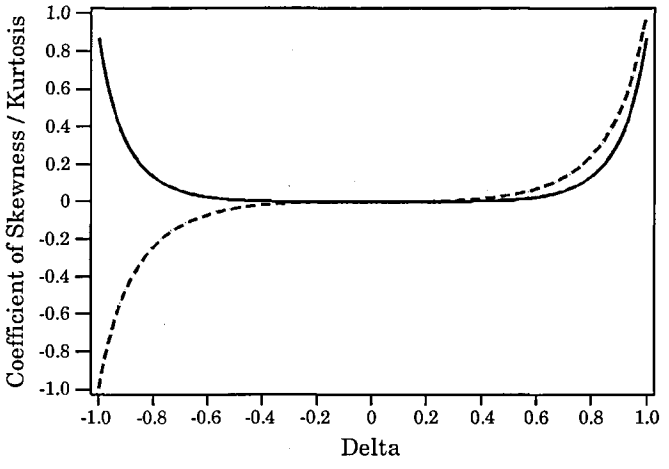

Figure 1.2 Coefficients of skewness (- - - - -) and kurtosis (———) as functions of the parameter δ.

Using (1.2.4), the first four moments and variance of Y_D are:

$$\begin{aligned}
&E(Y_{\text{D}}) = \xi + b\eta\delta, \quad E(Y_{\text{D}}^2) = \xi^2 + 2b\xi\eta\delta + \eta^2, \\
&E(Y_{\text{D}}^3) = \xi^3 + 3b\xi^2\eta\delta + 3\xi\eta^2 + 3b\eta^3\delta - b\eta^3\delta^3, \\
&E(Y_{\text{D}}^4) = \xi^4 + 4b\xi^3\eta\delta + 6\xi^2\eta^2 + 4b\xi\eta^3\delta(3-\delta^2) + 3\eta^4, \\
&\text{var}(Y_{\text{D}}) = \eta^2(1-b^2\delta^2).
\end{aligned} \tag{1.2.6}$$

The coefficients of skewness and kurtosis for Y_D are the same as those for X, namely,

$$\gamma_1 = \frac{b\delta^3(2b^2-1)}{(1-b^2\delta^2)^{3/2}} \in (-0.99527, 0.99527),$$

$$\gamma_2 = \frac{2(\pi-3)b^4\delta^4}{(1-b^2\delta^2)^2} \in [0, 0.86918).$$

In Figure 1.2 we represent these two measures as functions of δ. For reference purposes, we note that λ-values of 2, 5 and 20 correspond to δ-values of 0.8944, 0.9806 and 0.9988, respectively.

1.3 The Extended General Skew-normal Class

In an attempt to widen the ranges of skewness and kurtosis of the class, Azzalini (1985) proposed a further extension of it, introducing a shape parameter, $\zeta \in (-\infty, \infty)$. The random variable W_D is said to have an extended general skew-normal distribution if its pdf is given by

$$f(w;\xi,\eta,\lambda,\zeta) = \phi\left(\frac{w-\xi}{\eta}\right)\Phi\left\{\lambda\left(\frac{w-\xi}{\eta}\right)+\zeta\right\}\Big/\eta\,\Phi\left\{\zeta(1+\lambda^2)^{-1/2}\right\}, \qquad (1.3.1)$$

where $-\infty < w < \infty$. We denote the fact by $W_D \sim \mathrm{SNE}_D(\xi,\eta,\lambda,\zeta)$.

Henze (1986) provided a representation of an extended skew-normal random variable in terms of a normalized linear combination of a standard normal random variable and a truncated standard normal random variable, and derived expressions for the distribution's moments. Arnold & Beaver (2000) give the following construction which leads to a model described by (1.3.1). Suppose W and U are independent and identically distributed standard normal random variables. Then the standardized version of (1.3.1), i.e. with $\xi = 0$ and $\eta = 1$, is the conditional density of W given that $\zeta + \lambda W > U$.

Although the extended skew-normal distribution was developed independently by Azzalini (1985), its first appearance can be traced to Birnbaum (1950). The distribution had also been proposed previously by O'Hagan & Leonard (1976) as a potential skew prior when there is uncertainty about an inequality constraint in the bayesian estimation of a normal location parameter. Given these precedences, and those referred to in Section 1.2.1, there is clearly an issue as to the intellectual

"ownership" of the skew-normal distribution. On the one hand, the common reference to it as "Azzalini's skew-normal distribution" is a popularism which fails to convey the distribution's deeper historical roots. Nevertheless, the term does highlight the fundamental contribution of Azzalini and his co-authors towards the characterization, extension and dissemination of the distribution. Implicitly, it also suggests the potential existence of other skew-normal distributions (see Section 1.5.2).

Azzalini (1985) obtained the moment generating function of W_D and quoted the ranges of γ_1 and γ_2 as being approximately (–1.2, 1.2) and (0, 2), respectively. Arnold & Beaver (2002) discuss multivariate extensions of the class.

1.4 Inference for the General Skew-normal Distribution

Classical moment and likelihood based inference for problems concerning the general skew-normal distribution have been addressed in the literature by: Azzalini (1985), Salvan (1986), Arnold *et al.* (1993), Chiogna (1997), Azzalini & Capitanio (1999) and Pewsey (2000a). Liseo (1990) and Mukhopadhyay & Vidakovic (1995) consider inference from a bayesian perspective. In this section we discuss moment and likelihood based inference, the treatment given following closely that published in Pewsey (2000a).

1.4.1 Inference for the Direct Parametrization

1.4.1.1 Moment Based Inference

Inference based on the method of moments for the direct parametrization of the general skew-normal distribution has been considered in Arnold *et al.* (1993) and Pewsey (2000a). Proceeding as in Pewsey (2000a), let $\underline{y} = (y_1, ..., y_n)$ denote a random sample of n observations from a $\mathrm{SN}_D(\xi, \eta, \lambda)$ distribution, with sample moments $m_1 = 0, m_2 = s^2, ...$ about the mean. We denote the moment estimates of ξ, η and δ by $\tilde{\xi}$, $\tilde{\eta}$, $\tilde{\delta}$, respectively. For simplicity, consider the studentized sample $\underline{y}_s = (y_{s1}, ..., y_{sn})$ where $y_{si} = (y_i - \bar{y})/s$, $i = 1, ..., n$; this is a sample from a $\mathrm{SN}_D(\xi_s, \eta_s, \lambda)$ distribution with $\xi_s = (\xi - \bar{y})/s$ and $\eta_s = \eta/s$. Equating the first three sample moments of the studentized data to their population counterparts from (1.2.6), the method of moments (MM) estimates of ξ_s, η_s and δ are

$$\tilde{\xi}_s = -cm_3^{1/3}/s, \quad \tilde{\eta}_s = \left(1+\tilde{\xi}_s^2\right)^{1/2} \quad \text{and} \quad \tilde{\delta} = -\tilde{\xi}_s/b\tilde{\eta}_s, \tag{1.4.1}$$

where $c = \{2/(4-\pi)\}^{1/3}$. Then, $\tilde{\lambda} = \tilde{\delta}/\left(1-\tilde{\delta}^2\right)^{1/2}$, provided $|\tilde{\delta}|<1$. Otherwise, $\tilde{\delta}$ is out of range as an estimate of δ and we refer to it as being 'inadmissible'. For such estimates, $\tilde{\lambda}$ is undefined. The MM estimates of ξ and η can be recovered using

$$\tilde{\xi} = \bar{y} + s\tilde{\xi}_s \quad \text{and} \quad \tilde{\eta} = s\tilde{\eta}_s \tag{1.4.2}$$

This approach is equivalent to, although arguably simpler than, that given in Arnold *et al.* (1993).

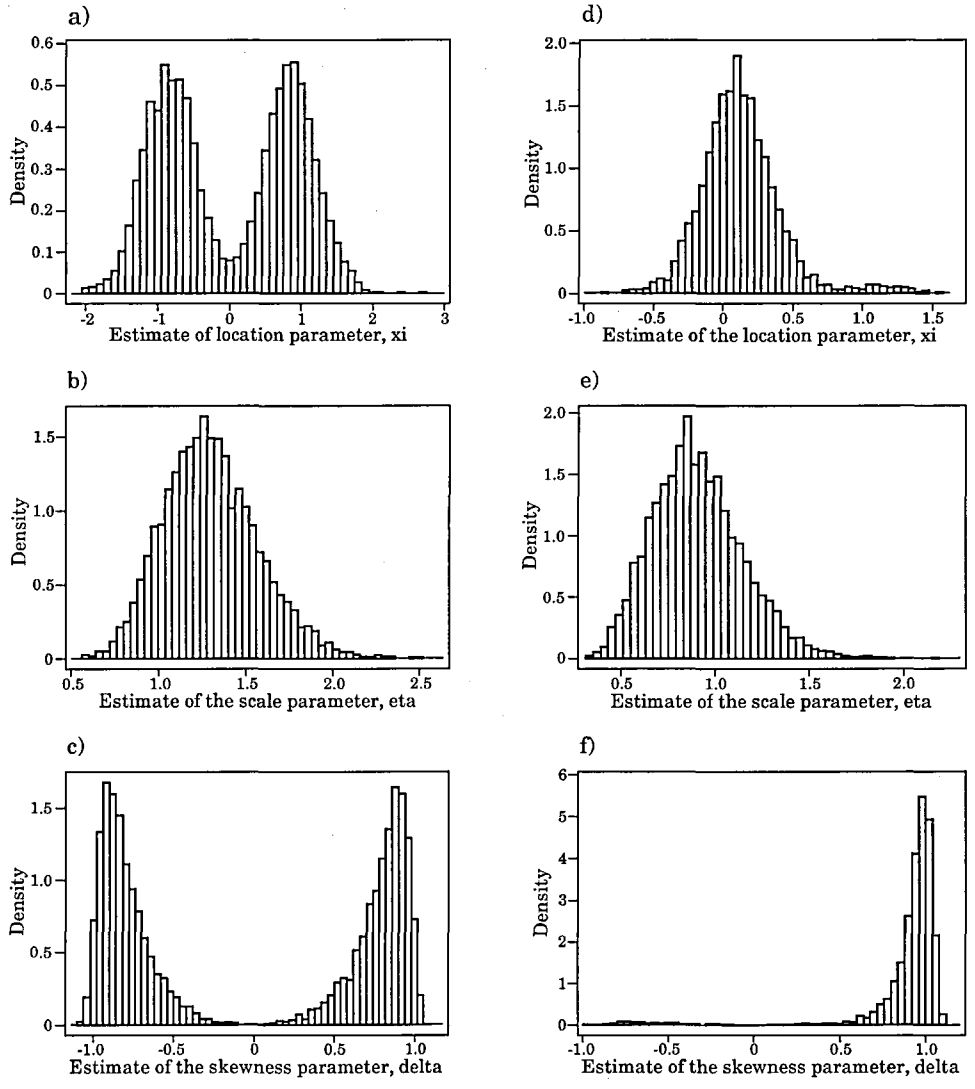

Figure 1.3 Empirical sampling distributions of the method of moments estimates of ξ, η and δ obtained from 5000 simulated samples of size 20 from the $SN_D(0,1,\lambda)$ distribution: a), b), c) $\lambda = 0$; d), e), f) $\lambda = 20$.

As mentioned above, a major deficiency of MM estimation is that the estimate of δ can be inadmissible, i.e. greater than, or equal to, 1 in absolute value. Such estimates occur with greater frequency as the skewness parameter of the underlying skew-normal distribution tends to $\pm\infty$. Faced with an inadmissible estimate, we interpret it as indicating that the underlying distribution is half-normal or negative half-normal, depending on its sign. Inference for the general half-normal distribution is considered in Chapter 2.

In terms of subsequent inference concerning the parameters, a further problematic feature of MM estimation for the direct parametrization is that the sampling distributions of the estimates of the location and skewness parameters are often bimodal, the problem being particularly acute when the underlying distribution is normal. For this particular case, $\bar{y}$ and s are the MM and maximum likelihood (ML) estimates of the location and scale parameters. However, if we use (1.4.1) and (1.4.2) to estimate the parameters of an assumed skew-normal distribution, when the data come from a normal population, we will tend to over-estimate η and over- or under-estimate ξ and δ, depending on the sign of $\tilde{\xi}_s$.

Figures 1.3 and 1.4 provide representations of the empirical sampling distributions of the method of moments estimates for simulated samples from the symmetrical standard normal distribution and the highly skewed $SN_D(0, 1, 20)$ distribution. We have chosen to represent the skewness of the distribution using δ rather than λ so as to avoid the complications of infinite or undefined estimates of the latter. In these two figures, values of $|\tilde{\delta}|$ in excess of 1 are inadmissible. As can be appreciated, as the sample size and magnitude of the skewness parameter increase the sampling distributions of all three estimates tend to unimodal distributions. However, these sampling distributions cannot realistically be considered as being normal, even for sample sizes as large as 500. For the samples of size 20, the sampling distribution of $\tilde{\delta}$ is still bimodal even for data drawn from a highly skewed distribution. We also note that, for the samples from the $SN_D(0, 1, 20)$ distribution, the proportion of inadmissible δ-estimates increases spectacularly with increasing sample size. Increasing sample size does nothing to alleviate the problems associated with the form of the sampling distributions of the estimates for data drawn from the standard normal distribution.

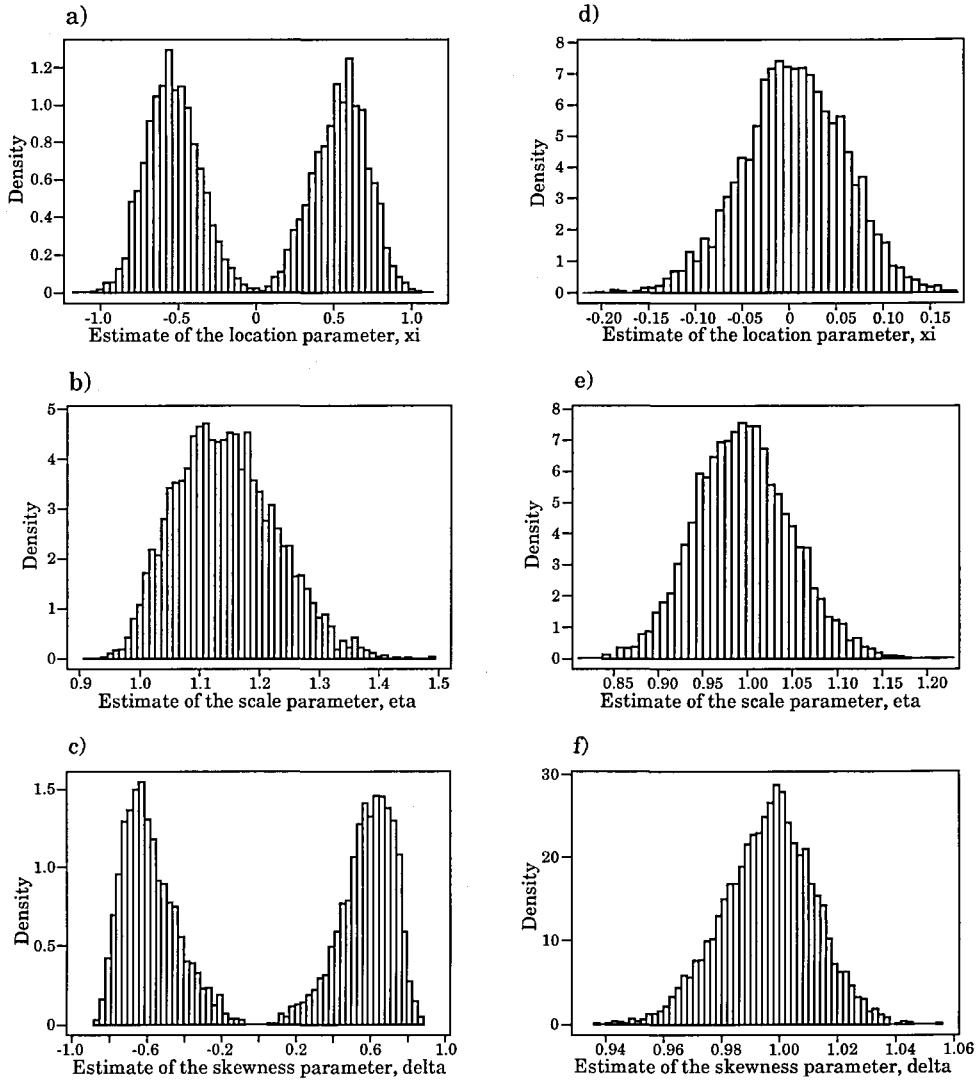

Figure 1.4 Empirical sampling distributions of the method of moments estimates of ξ, η and δ obtained from 5000 simulated samples of size 500 from the $SN_D(0,1,\lambda)$ distribution: a), b), c) $\lambda = 0$; d), e), f) $\lambda = 20$.

Whilst some of the problems referred to above regarding the sampling properties of the MM estimators were also mentioned in Arnold *et al.* (1993), these authors failed to identify the problems we have referred to when the underlying distribution is normal. We consider this oversight to be due to the limited scope of their reported simulation study which only included pseudo-random variates generated from a $SN_D(10, 5, 2)$ distribution. Given that MM estimation can be carried out using studentization in conjunction with (1.4.1) and (1.4.2), the choices of $\xi = 10$ and $\eta = 5$ in their simulations are somewhat irrelevant. As we have commented in Section 1.2.2, a λ-value of 2 corresponds to a density which is moderately asymmetric.

1.4.1.2 Likelihood Based Inference

Properties of the likelihood function, and inference based upon it, for the direct parametrization of the skew-normal distribution have been considered by Azzalini (1985), Liseo (1990), Arnold *et al.* (1993), Azzalini & Dalla Valle (1996), Chiogna (1997), Azzalini & Capitanio (1999) and Pewsey (2000a). Three equations satisfied by the ML estimates corresponding to a random sample, $\underline{y}$, from the $\mathrm{SN}_{\mathrm{D}}(\xi,\eta,\lambda)$ distribution are given in Arnold *et al.* (1993). For our purposes, the most relevant of the three is

$$\hat{\eta}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\xi}\right)^2, \tag{1.4.3}$$

a constraint on the ML estimates $\hat{\xi}$ and $\hat{\eta}$ which had previously been identified in Azzalini (1985). In the same paper, Azzalini gives the Fisher information matrix for (ξ,η,λ) and notes that it is singular for $\lambda = 0$. Arnold *et al.* (1993) highlight the fact that $\lambda = 0$, $\xi = \bar{y}$ and $\eta^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ is always a solution to the likelihood equations, although, in general, this solution does not give the ML estimates. Azzalini (1985) gives an algorithm for finding the ML estimates based on the profile log-likelihood of λ. However, as pointed out by Arnold *et al.* (1993), the quoted algorithm is incorrect as the constraint (1.4.3) does not apply to the profile log-likelihood. Azzalini (1985) observes that the profile log-likelihood of λ always has a stationary point at $\lambda = 0$. Arnold *et al.* (1993) also consider profile likelihood methods and quote numerical results which indicate that this stationary point in the profile log-likelihood corresponds to a saddlepoint on the likelihood surface. Chiogna (1997) proves that $\lambda = 0$ is a point of inflexion for the profile log-likelihood of λ.

To illustrate some of the problems associated with the likelihood for the direct parametrization we use a form of the full log-likelihood function incorporating the constraint (1.4.3) and studentization of the original data as described in Section 1.4.1.1. For the studentized data the constraint (1.4.3) becomes $\hat{\eta}_s = \left(1+\hat{\xi}_s^2\right)^{1/2}$, an analogous relation to that given by the middle constraint of (1.4.1) for the equivalent MM estimates. Using this constraint for the studentized data along with the density (1.2.5), maximum likelihood estimation reduces to finding, using

numerical optimization, those values of ξ_s and λ which maximize the constrained log-likelihood

$$l(\xi_s, \lambda; \underline{y}_s) = -\frac{n}{2}\log_e(1+\xi_s^2) + \sum_{i=1}^{n}\log_e \Phi\left\{\frac{\lambda(y_{si}-\xi_s)}{(1+\xi_s^2)^{1/2}}\right\}. \qquad (1.4.4)$$

In theory, the ML estimates of the location and scale parameters for the original, non-studentized, data can be obtained by substituting the ML estimates $\hat{\xi}_s$ and $\hat{\eta}_s$, the latter calculated using (1.4.3), in place of their MM counterparts in (1.4.2).

Table 1.1 The frontier data of Azzalini & Capitanio (1999).

0.1169	1.7311	0.0144	0.4881	2.3877	-0.0853	0.0522	0.7226	0.8718	3.0415
0.6363	1.3590	1.5958	0.4567	0.9885	1.7001	1.0380	1.0195	0.3528	1.4249
0.3170	1.3276	-0.1032	1.1568	0.0699	1.6802	0.2470	0.4147	1.6882	0.7256
1.3568	1.1091	0.0500	2.2886	1.4985	2.7261	1.8443	-0.0687	0.9441	0.6872
0.5258	0.4743	0.4240	0.7349	0.4428	0.1880	0.4642	0.2786	0.2742	0.5678

Log-likelihood functions defined by (1.4.4) generally contain a long narrow ridge which is often curved. Arnold *et al.* (1993) refer to similar features in the two dimensional profile log-likelihood functions associated with the direct parametrization. This type of feature can cause problems for many optimization methods, as such techniques generally perform well for surfaces with contours which are close to symmetrical ellipsoids; see e.g. Lindsey (1996, p. 109). For surfaces containing narrow ridges, certain iterative methods may fail to converge, let alone converge to the required global maximum. This is a known feature, for instance, of gradient-based iterative techniques. Given the generally non-elliptical form of the likelihood surface under the direct parametrization, convergence, when it occurs, can be slow. The constrained log-likelihood surface defined by (1.4.4) for the so-called "frontier data" of Azzalini & Capitanio (1999) exhibits such a curved ridge feature, as can be appreciated from Figure 1.5. In this figure we have represented (1.4.4) in terms of δ, rather than λ, so as to avoid the complications associated with the range of the latter. The frontier data, reproduced in Table 1.1, consist of 50 pseudo-random variates simulated from the $SN_D(0, 1, 5)$ distribution.

Unlike its method of moments counterpart, a maximum likelihood estimate of a parameter is, by definition, always within range, i.e. 'admissible'. However, particularly for small sized samples, and generally for samples drawn from skew-normal distributions with moderate to large values of λ, the global maximum of the

log-likelihood can occur on the boundary of the parameter space. In such cases, the maximum likelihood estimate of the skewness parameter λ is infinite. The frontier data provide a case in point. Indeed, the word "frontier" used in their description presumably refers to the fact that the maximum likelihood estimate of the skewness parameter occurs on the boundary, or "frontier", of the parameter space. As we can see from Figure 1.5, the prominent narrow ridge winds its way out to the global maximum corresponding to a value on the $\delta = 1$ (or, equivalently, $\lambda = +\infty$) boundary. As an isolated case, this behaviour is somewhat surprising as the sample distribution of the frontier data displays no obviously pathological features. For instance, the sample coefficient of skewness is 0.9022, well below the theoretical maximum for a skew-normal random variable of 0.9953. The corresponding value for the population coefficient of skewness for the underlying $SN_D(0, 1, 5)$ distribution is 0.8510. As Cox (1992) and Lindsey (1996, p. 81) note, boundary estimates of parameters have a clear interpretation. Formally, a boundary maximum likelihood estimate indicates that the set of plausible models lies to only one side of the most likely one. Thus, here, we interpret a boundary estimate of λ as indicating that a half-normal distribution is the most likely generating mechanism for the data. Again, see Chapter 2 for a treatment of inference for the general half-normal distribution.

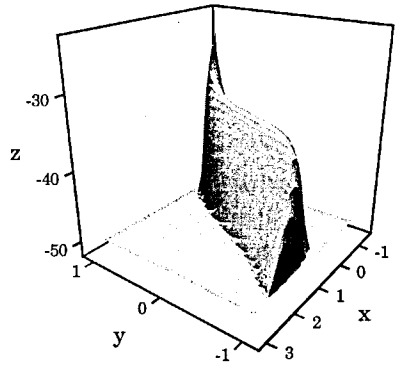

Figure 1.5 Constrained log-likelihood surface under the direct parametrization, truncated at $z = -50$, for the studentized values of the frontier data: $x = \xi_s$, $y = \delta$, z = value of (1.4.4).

If operating with infinite values of a parameter estimate is thought messy then we can always reparametrize. For instance, here we could choose to work with $\delta \in (-1,1)$ rather than $\lambda \in (-\infty,\infty)$. Whilst working with, for instance, δ rather than

λ circumvents the problem of infinite point estimates, boundary estimates on the λ scale are still nevertheless transformed to boundary estimates on the δ scale. And boundary values for point estimates, under whatever parametrization, are problematic in that the regularity conditions underpinning standard, asymptotic theory based, likelihood methods of inference do not apply on the boundary of the parameter space.

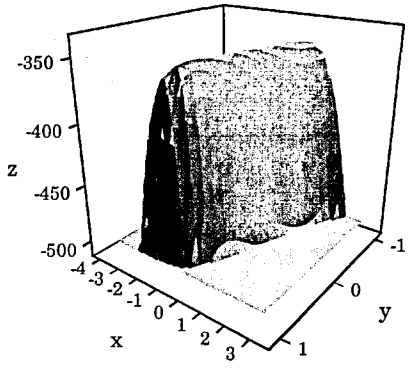

Figure 1.6 Constrained log-likelihood surface under the direct parametrization, truncated at $z = -500$, for the studentized values of a simulated sample of size 500 from the $N(0,1) \equiv SN_D(0,1,0)$ distribution: $x = \xi_s$, $y = \delta$, $z =$ value of (1.4.4).

Pewsey (2000a) traces the singularity of the Fisher information matrix for $\lambda = 0$ to the parameter redundancy of the parametrization for the normal case, a fact easily identified using the results of Catchpole & Morgan (1997). In the latter, an exponential-family model is identified as being parameter redundant if the mean can be expressed using a reduced number of parameters. From (1.2.6), $E(Y_D)$ is a function of all three parameters, whereas for the normal case it is just ξ. The singularity of the information matrix then follows from Remark 4 of Catchpole & Morgan (1997). According to Theorem 2 of the same paper, the likelihood surface for the normal case must contain a completely flat ridge. Thus, if we were to attempt to maximize the log-likelihood for this parametrization using numerical techniques, the results obtained could be highly misleading as for this case no unique solution exists. In Figure 1.6 we plot (1.4.4) for a simulated sample of size 500 from the $N(0, 1) \equiv SN_D(0, 1, 0)$ distribution. As in Figure 1.5, we plot (1.4.4) as a function of δ, and truncate the surface so as to highlight its main features. We see that even for such a large sample size, the surface is completely flat across a

relatively wide range of δ-values, indicating that the data provide little or no information about the parameters. We also note that the dominant ridge in Figure 1.6 is not orthogonal to either parameter axis and so neither the skewness nor the location parameter have unique maximum likelihood estimates. These consequences, and our findings for MM estimation, rule the direct parametrization out as a general basis from which to conduct estimation.

1.4.2 Inference for the Centred Parametrization

Having identified the singularity problem associated with ML estimation for the direct parametrization, Azzalini (1985) introduced the "centred" parametrization, (μ, σ, γ_1). He defined a skew-normal random variable Y_C with $E(Y_C) = \mu$ and $\text{var}(Y_C) = \sigma^2$ by

$$Y_C = \mu + \sigma \frac{\{X - E(X)\}}{\{\text{var}(X)\}^{1/2}}, \quad -\infty < \mu < \infty, \sigma > 0,$$

where X is a $\text{SN}(\lambda)$ random variable. In this notation, we use the subindex C to highlight the role of the centred parametrization. The parameter γ_1 is the coefficient of skewness of X, and hence also that of Y_C. We denote the distribution of Y_C under this parametrization by $\text{SN}_C(\mu, \sigma, \gamma_1)$. As $E(Y_C) = \mu$, this parametrization is clearly not parameter redundant for the normal case. Expressions for the direct parameters in terms of the centred ones are:

$$\begin{aligned} \xi &= \mu - c\gamma_1^{1/3}\sigma, \\ \eta &= \sigma\left(1 + c^2\gamma_1^{2/3}\right)^{1/2}, \\ \lambda &= \frac{c\gamma_1^{1/3}}{\left\{b^2 + c^2\left(b^2 - 1\right)\gamma_1^{2/3}\right\}^{1/2}}, \end{aligned} \tag{1.4.5}$$

where, as before, $b = (2/\pi)^{1/2}$ and $c = \{2/(4-\pi)\}^{1/3}$. The density of Y_C under this parametrization is given by

$$f(y;\mu,\sigma,\gamma_1)=\frac{2}{\sigma\left(1+c^2\gamma_1^{2/3}\right)^{1/2}}\phi\left[\frac{1}{\left(1+c^2\gamma_1^{2/3}\right)^{1/2}}\left\{\left(\frac{y-\mu}{\sigma}\right)+c\gamma_1^{1/3}\right\}\right]$$
$$\times\Phi\left[\frac{c\gamma_1^{1/3}}{\left\{b^2+c^2\left(b^2-1\right)\gamma_1^{2/3}\right\}^{1/2}\left(1+c^2\gamma_1^{2/3}\right)^{1/2}}\left\{\left(\frac{y-\mu}{\sigma}\right)+c\gamma_1^{1/3}\right\}\right] \qquad (1.4.6)$$

$$\left(-\infty<y<\infty;-\infty<\mu<\infty;\ \sigma>0;-0.99527<\gamma_1<0.99527\right).$$

1.4.2.1 Moment Based Inference

Moment based inference for the centred parametrization has only been considered in Pewsey (2000a). As,

$$E(Y_C)=\mu,\ E(Y_C^2)=\mu^2+\sigma^2 \text{ and } E(Y_C^3)=\mu^3+3\mu\sigma^2+\sigma^3\gamma_1, \qquad (1.4.7)$$

the MM estimates of the centred parameters are the usual ones, namely

$$\tilde{\mu}=\bar{y},\ \tilde{\sigma}=s,\ \tilde{\gamma}_1=g_1=m_3/s^3.$$

Prior to giving a general result for the asymptotic distribution of $(\bar{y}, s, g_1)$ in Theorem 1.1, we provide the details of a lemma quoted by Mardia (1972, p. 111) which summarizes a standard means of deriving large-sample approximations to sampling distributions based on Taylor expansion.

Lemma 1.2 *Let the p-dimensional statistic* $(T_1,...,T_p)^T$ *have joint distribution which is asymptotically* $N(\underline{\theta},\underline{\Sigma})$ *with* $\underline{\theta}=(\theta_1,...,\theta_p)^T$ *and* $\underline{\Sigma}=(\sigma_{ij})$, *where* $E(T_i)=\theta_i$, *and* $\mathrm{var}(T_i)=\sigma_{ii}$ *and* $\mathrm{cov}(T_i,T_j)=\sigma_{ij}$ *are of order* n^{-1}. *Further, let* $h_1,...,h_q$ *be differentiable functions of* $T_1,...,T_p$. *Then*

$$E(h_k)=h_k(\theta_1,...,\theta_p)+\tfrac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}h_k^{(ij)}\sigma_{ij}+O\left(n^{-3/2}\right),$$
$$\mathrm{var}(h_k)=\sum_{i=1}^{p}\sum_{j=1}^{p}h_k^{(i)}h_k^{(j)}\sigma_{ij}+O\left(n^{-3/2}\right), \qquad (1.4.8)$$
$$\mathrm{cov}(h_k,h_l)=\sum_{i=1}^{p}\sum_{j=1}^{p}h_k^{(i)}h_l^{(j)}\sigma_{ij}+O\left(n^{-3/2}\right),$$

where $k, l = 1,...,q$ *and* $h_k^{(i)}=\partial h_k/\partial\theta_i$, $h_k^{(ij)}=\partial^2 h_k/\partial\theta_i\,\partial\theta_j$ *and* $h_k\equiv h_k(\underline{\theta})$. *The joint distribution of* $(h_1,...,h_q)$ *is asymptotically normal if the leading terms of the* $E(h_k)$ *are finite and those for the* $\mathrm{var}(h_k)$ *are of the form* c_k/n *with* $c_k>0$.

Theorem 1.1 *Let $Y_1,...,Y_n$ be n independently and identically distributed random variables from a distribution for which the first six central moments are finite. Let μ, σ^2 and γ_1 denote the mean, variance and coefficient of skewness of the distribution, respectively. Let $\beta_2 = \mu_4/\mu_2^2$, $\beta_3 = \mu_5/\mu_2^{5/2}$ and $\beta_4 = \mu_6/\mu_2^3$, where μ_r denotes the r^{th} central moment about the mean. Then the asymptotic joint distribution of $(\bar{y}, s, g_1)$ is trivariate normal with*

$$\begin{aligned} &E(\bar{y}) = \mu, \\ &E(s) = \sigma\{1-(3+\beta_2)/8n\} + O(n^{-3/2}), \\ &E(g_1) = \gamma_1 + 3\{\gamma_1(7+5\beta_2) - 4\beta_3\}/8n + O(n^{-3/2}), \end{aligned} \tag{1.4.9}$$

$$\begin{aligned} &\mathrm{var}(\bar{y}) = \sigma^2/n, \\ &\mathrm{var}(s) = \sigma^2(\beta_2 - 1)/4n + O(n^{-3/2}), \\ &\mathrm{var}(g_1) = \{9 - 6\beta_2 - 3\gamma_1\beta_3 + \beta_4 + \gamma_1^2(35+9\beta_2)/4\}/n + O(n^{-3/2}), \end{aligned} \tag{1.4.10}$$

$$\begin{aligned} &\mathrm{cov}(\bar{y}, s) = \sigma^2\gamma_1/2n + O(n^{-3/2}), \\ &\mathrm{cov}(\bar{y}, g_1) = \sigma(\beta_2 - 3 - 3\gamma_1^2/2)/n + O(n^{-3/2}), \\ &\mathrm{cov}(s, g_1) = \sigma\{2\beta_3 - \gamma_1(5+3\beta_2)\}/4n + O(n^{-3/2}), \end{aligned} \tag{1.4.11}$$

if the terms of $O(n)$ in the variances of s and g_1 are positive.

Proof Let $T_1 = m_1'$, $T_2 = m_2'$ and $T_3 = m_3'$ where $m_k' = \frac{1}{n}\sum_{i=1}^{n} Y_i^k$. Then, $\bar{y} = T_1$, $s = (T_2 - T_1^2)^{1/2}$ and $g_1 = \frac{T_3 - 3T_2T_1 + 2T_1^3}{(T_2 - T_1^2)^{3/2}}$, differentiable functions of T_1, T_2 and T_3. The non-central moments can be expressed as,

$$\begin{aligned} &E(Y) = \mu, \\ &E(Y^2) = \mu^2 + \sigma^2, \\ &E(Y^3) = \mu^3 + 3\mu\sigma^2 + \sigma^3\gamma_1, \\ &E(Y^4) = \mu^4 + 6\mu^2\sigma^2 + 4\mu\sigma^3\gamma_1 + \sigma^4\beta_2, \\ &E(Y^5) = \mu^5 + 10\mu^3\sigma^2 + 10\mu^2\sigma^3\gamma_1 + 5\mu\sigma^4\beta_2 + \sigma^5\beta_3, \\ &E(Y^6) = \mu^6 + 15\mu^4\sigma^2 + 20\mu^3\sigma^3\gamma_1 + 15\mu^2\sigma^4\beta_2 + 6\mu\sigma^5\beta_3 + \sigma^6\beta_4. \end{aligned} \tag{1.4.12}$$

Appealing to the central limit theorem, the asymptotic distribution of $(T_1, T_2, T_3)^T$ is $N(\underline{\theta}, \underline{\Sigma})$ where $\underline{\theta} = (\theta_1, \theta_2, \theta_3)^T = \{E(Y), E(Y^2), E(Y^3)\}^T$ and $\underline{\Sigma} = (\sigma_{ij})$, with

$$
\begin{aligned}
\sigma_{11} &= \mathrm{var}(T_1) = \frac{\sigma^2}{n}, \\
\sigma_{12} &= \sigma_{21} = \mathrm{cov}(T_1, T_2) = \frac{\sigma^2}{n}(2\mu + \sigma\gamma_1), \\
\sigma_{13} &= \sigma_{31} = \mathrm{cov}(T_1, T_3) = \frac{\sigma^2}{n}(3\mu^2 + 3\mu\sigma\gamma_1 + \sigma^2\beta_2), \\
\sigma_{22} &= \mathrm{var}(T_2) = \frac{\sigma^2}{n}\{4\mu^2 + 4\mu\sigma\gamma_1 + \sigma^2(\beta_2 - 1)\}, \\
\sigma_{23} &= \sigma_{32} = \mathrm{cov}(T_2, T_3) \\
&= \frac{\sigma^2}{n}\{6\mu^3 + 9\mu^2\sigma\gamma_1 + \mu\sigma^2(5\beta_2 - 3) + \sigma^3(\beta_3 - \gamma_1)\}, \\
\sigma_{33} &= \mathrm{var}(T_3) = \frac{\sigma^2}{n}\{9\mu^2(\mu^2 - \sigma^2) + 6\mu\sigma(3\mu^2 - \sigma^2)\gamma_1 - \sigma^4\gamma_1^2 \\
&\qquad + 15\mu^2\sigma^2\beta_2 + 6\mu\sigma^3\beta_3 + \sigma^4\beta_4\}.
\end{aligned}
\tag{1.4.13}
$$

Now apply Lemma 1.2 with $h_1 = \theta_1$, $h_2 = (\theta_2 - \theta_1^2)^{1/2}$ and $h_3 = \dfrac{\theta_3 - 3\theta_1\theta_2 + 2\theta_1^3}{(\theta_2 - \theta_1^2)^{3/2}}$.

The non-zero first- and second-order partial derivatives of these functions, expressed in terms of μ, σ^2 and γ_1, are:

$$
\begin{aligned}
h_1^{(1)} &= 1, \\
h_2^{(1)} &= \frac{-\mu}{\sigma}, \quad h_2^{(2)} = \frac{1}{2\sigma}, \\
h_2^{(11)} &= \frac{-(\mu^2 + \sigma^2)}{\sigma^3}, \quad h_2^{(12)} = \frac{\mu}{2\sigma^3}, \quad h_2^{(22)} = \frac{-1}{4\sigma^3}, \\
h_3^{(1)} &= \frac{3(\mu^2 + \mu\sigma\gamma_1 - \sigma^2)}{\sigma^3}, \quad h_3^{(2)} = \frac{-3(2\mu + \sigma\gamma_1)}{2\sigma^3}, \quad h_3^{(3)} = \frac{1}{\sigma^3},
\end{aligned}
$$

$$h_3^{(11)} = \frac{3\{2\mu(3\mu^2 - \sigma^2) + \sigma\gamma_1(5\mu^5 + \sigma^2)\}}{\sigma^5},$$

$$h_3^{(12)} = \frac{3\{-9\mu^2 - 5\mu\sigma\gamma_1 + \sigma^2\}}{2\sigma^5}, \quad h_3^{(13)} = \frac{3\mu}{\sigma^5},$$

$$h_3^{(22)} = \frac{3(12\mu + 5\sigma\gamma_1)}{4\sigma^5}, \quad h_3^{(23)} = \frac{-3}{2\sigma^5}.$$

Substituting these partial derivatives, and the variances and covariances given by (1.4.13), in (1.4.8) we obtain, after lengthy but basic algebraic simplification, the expectations, variances and covariances given in (1.4.9)-(1.4.11). Asymptotic normality follows from Lemma 1.2 due to the asymptotic normality of $(T_1, T_2, T_3)^T$, the forms of the leading terms for the expectations in (1.4.9) and the assumption that the terms of $O(n)$ in the variances of s and g_1 in (1.4.10) are positive.

We suspect that Taylor expansion may well have been used previously in the literature to derive equivalent results to those given in Theorem 1.1 for the joint distribution of $(\bar{y}, s, g_1)$. Nevertheless, we have been unable to track down any such previous work. We note that Gupta (1967) used Taylor expansion to derive an asymptotic result for the marginal distribution of g_1.

In order to apply Theorem 1.1 in the case of an underlying general skew-normal distribution, we note that if $Y_C \sim \mathrm{SN}_C(\mu, \sigma, \gamma_1)$,

$$\begin{aligned}
\beta_2 &= \gamma_2 + 3 = 3 + 2\tau^4(\pi - 3), \\
\beta_3 &= 10\gamma_1 + \tau^5(3\pi^2 - 40\pi + 96)/4, \\
\beta_4 &= 15\{1 + \tau^4(2\pi - 6)\} - \tau^6(9\pi^2 - 80\pi + 160)/2,
\end{aligned}$$

where $\tau = c\gamma_1^{1/3}$.

For the special case of an underlying normal distribution, the relevant sampling properties of all three individual estimators are well known. Nevertheless, in Figures 1.7 and 1.8 we represent the sampling distributions of the method of moments estimates for the same simulated samples used to generate Figures 1.3 and 1.4. We do this so as to provide a comparison with the sampling distributions of the maximum likelihood estimates presented later in Figures 1.13 and 1.14.

Simulation confirms that all of the results in (1.4.9)-(1.4.11), apart from that for $\mathrm{var}(g_1)$, provide very good approximations to the sampling properties of the

estimators for data drawn from skew-normal distributions, even when the sample size is as small as 20. Sample sizes of 50 or more are required before the expression for $\text{var}(g_1)$ gives a reasonable approximation. Moreover, it is known that the sampling distribution of g_1 tends to normality very slowly even for data from normal populations; see e.g. Pearson (1963) and D'Agostino (1970).

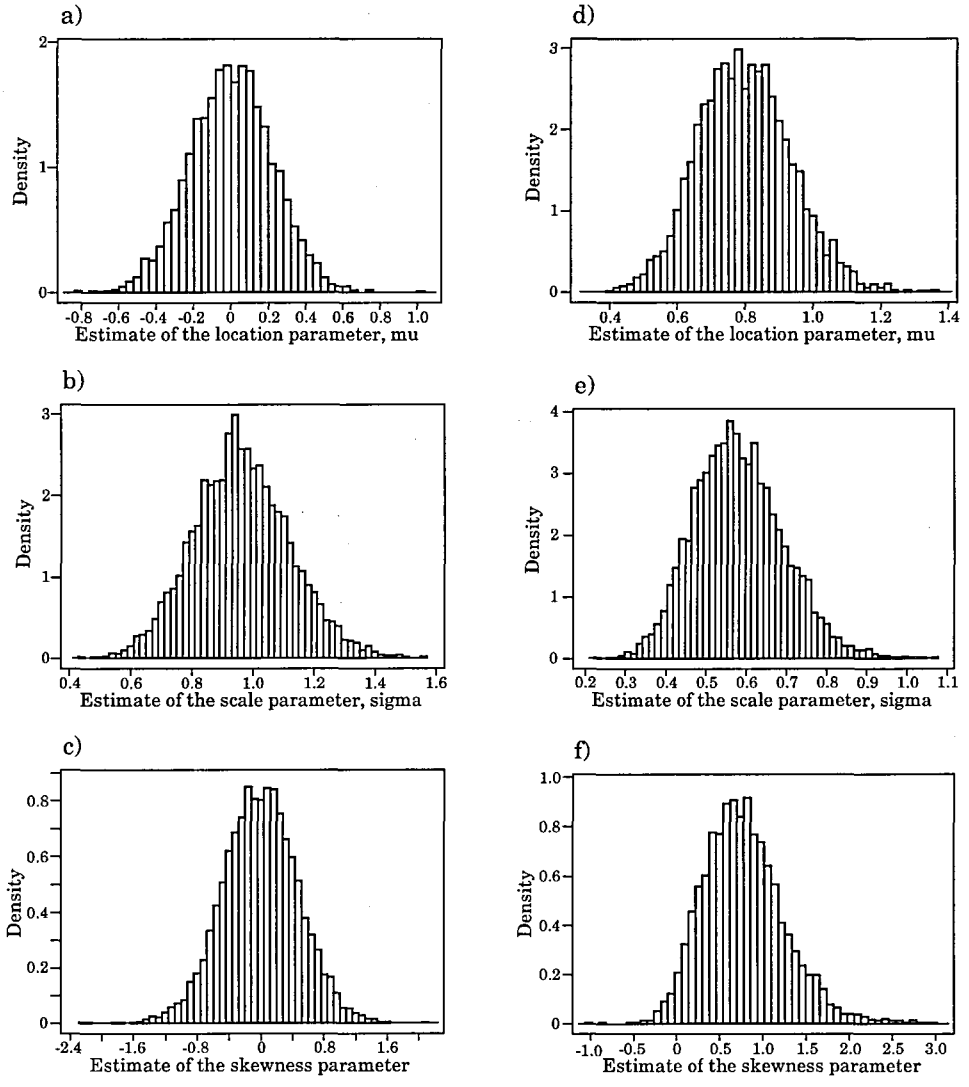

Figure 1.7 Empirical sampling distributions of the method of moments estimates of μ, σ and γ_1 obtained from the same 5000 simulated samples of size 20 used to produce Figure 1.3: a), b), c) $SN_D(0,1,0) \equiv SN_C(0,1,0)$; d), e), f) $SN_D(0,1,20) \equiv SN_C(0.7969, 0.6041, 0.9851)$.

For data from highly skewed cases of the skew-normal class the sampling distribution of g_1 is skewed, even for very large n, as can be appreciated from Figures 1.7 and 1.8. Should asymptotic theory be thought not to apply, inference for

γ_1 can be based upon computer intensive methods such as Monte Carlo significance testing and the bootstrap.

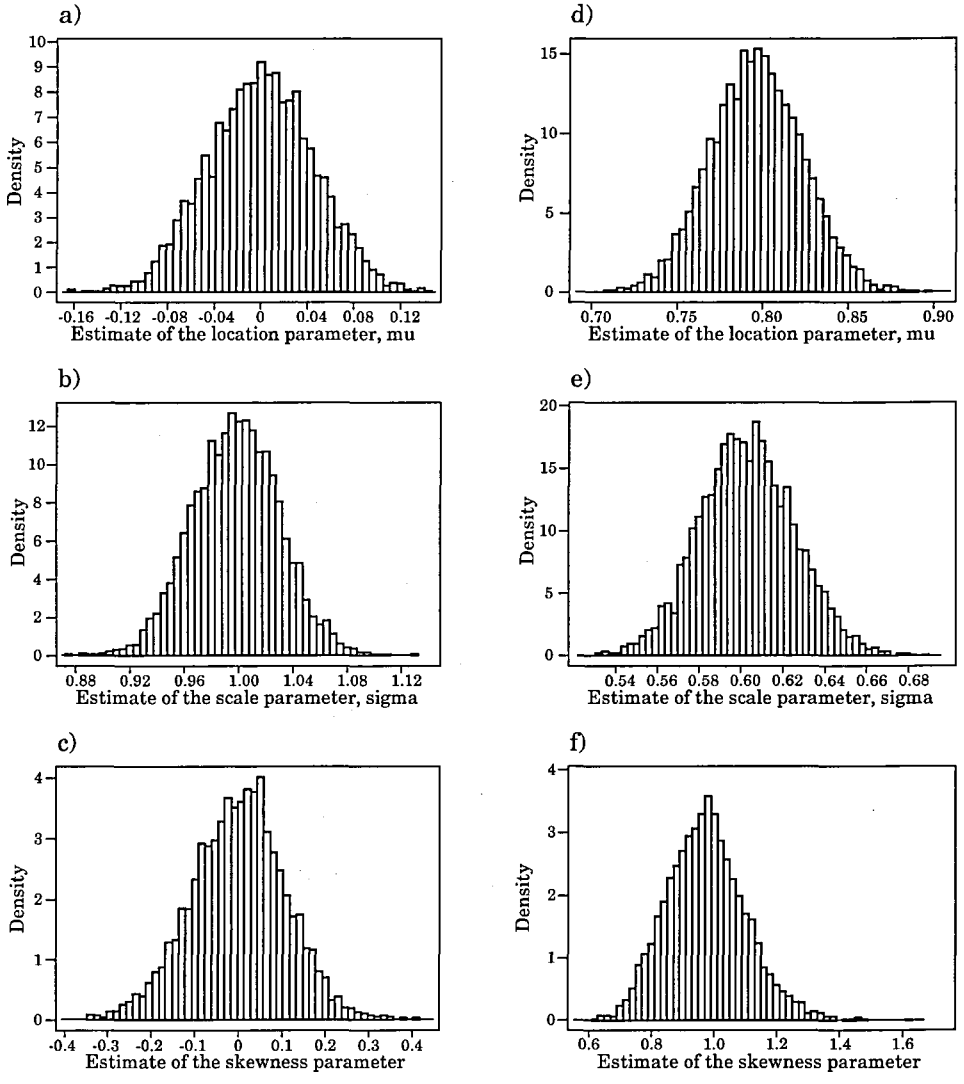

Figure 1.8 Empirical sampling distributions of the method of moments estimates of μ, σ and γ_1 obtained from the same 5000 simulated samples of size 500 used to produce Figure 1.4: a), b), c) $SN_D(0,1,0) \equiv SN_C(0,1,0)$; d), e), f) $SN_D(0,1,20) \equiv SN_C(0.7969, 0.6041, 0.9851)$.

As for the direct parametrization, a major problem associated with method of moments estimation for the centred parametrization is the occurrence of inadmissible estimates of the skewness parameter. Under the centred parametrization, inadmissible values of $\tilde{\gamma}_1 = g_1$ are those with absolute values in excess of 0.99527. Again, we interpret inadmissible estimates of γ_1 as indicating

that a half-normal distribution is the underlying generating mechanism. Approximately 3.7% of the estimates in Figure 1.7c, and none of the estimates in Figure 1.8c, are inadmissible. The corresponding percentages for the histograms in Figures 1.7f and 1.8f are 29 and 42, respectively.

We close this section by giving the following results for the frontier data. The theoretical values of the three parameters for a $SN_D(0, 1, 5)$ distribution are μ = 0.7824, σ = 0.6228 and γ_1 = 0.8510. The MM estimates of these parameters are $\tilde{\mu}$ = 0.8849, $\tilde{\sigma}$ = 0.7488 and $\tilde{\gamma}_1$ = 0.9022. These estimates are contrasted with the corresponding ML estimates in the following section.

1.4.2.2 Likelihood Based Inference

Work on issues associated with likelihood based inference for the centred parametrization of the skew-normal distribution has been published by Azzalini (1985), Azzalini & Dalla Valle (1996), Chiogna (1997), Azzalini & Capitanio (1999) and Pewsey (2000a). Considering theoretical results first, Azzalini (1985) gave the information matrix for the centred parameters. As proven by Chiogna (1997), the information matrix converges to $\mathrm{diag}\left(n/\sigma^2, 2n/\sigma^2, n/6\right)$ as $\gamma_1 \to 0$, not to its inverse as stated incorrectly in Azzalini (1985). The inverse of this diagonal matrix also corresponds to the asymptotic form of the covariance matrix for the method of moments estimators of the centred parameters as $\gamma_1 \to 0$, as can be established from (1.4.10) and (1.4.11). Chiogna (1997) shows that the derivative of the profile log-likelihood for γ_1 is finite and different from zero as $\gamma_1 \to 0$.

In the remainder of this section, we consider competing numerical approaches for identifying the maximum likelihood estimates of the centred parameters. Particular attention is given to the approach proposed in Pewsey (2000a) based on a constrained version of the full log-likelihood function analogous to that in (1.4.4).

Consider a random sample $\underline{y}$ from the $SN_C\left(\mu, \sigma, \gamma_1\right)$ distribution and its studentized counterpart $\underline{y}_s$ from the $SN_C\left(\mu_s, \sigma_s, \gamma_1\right)$ distribution. The constraint (1.4.3) on the ML estimates for the direct parametrization leads to the constraint $\hat{\sigma}_s = \left\{\hat{\mu}_s^2\left(1+\hat{\tau}^2\right)+1\right\}^{1/2} - \hat{\mu}_s\hat{\tau}$, where $\hat{\tau} = c\hat{\gamma}_1^{1/3}$. The ML estimates, $\hat{\mu}_s$ and $\hat{\gamma}_1$, are those values which maximize the constrained log-likelihood

$$l(\mu_s,\gamma_1;\underline{y}_s)=-n\log_e\left[\left\{\mu_s^2(1+\tau^2)+1\right\}^{1/2}-\mu_s\tau\right]-\frac{n}{2}\log_e(1+\tau^2)$$
$$+\sum_{i=1}^{n}\log_e\left\{\Phi\left(\left[\frac{(y_{si}-\mu_s)\tau}{\left\{\mu_s^2(1+\tau^2)+1\right\}^{1/2}-\mu_s\tau}+\tau^2\right]\Big/\left[\left\{b^2+\tau^2(b^2-1)\right\}(1+\tau^2)\right]^{1/2}\right)\right\}, \quad (1.4.14)$$

$-\infty<\mu_s<\infty, -0.99527<\gamma_1<0.99527$. The ML estimates of μ and σ are then given by $\hat{\mu}=\bar{y}+s\hat{\mu}_s$ and $\hat{\sigma}=s\hat{\sigma}_s$.

Although (1.4.14) is algebraically more complicated than (1.4.4), the surface defined by it is generally far better behaved. In Figure 1.9 we represent the surface obtained using (1.4.14) for the frontier data. Rather than having the winding ridge displayed in Figure 1.5, the main ridge in Figure 1.9 is close to being orthogonal to the μ_s axis. Nevertheless, the ridge still leads to a boundary estimate of the skewness parameter and thus the contours of the surface are clearly not elliptical. We also note that the reparametrization has removed the plateau feature of the ridge in the vicinity of $\lambda=\delta=\gamma_1=0$.

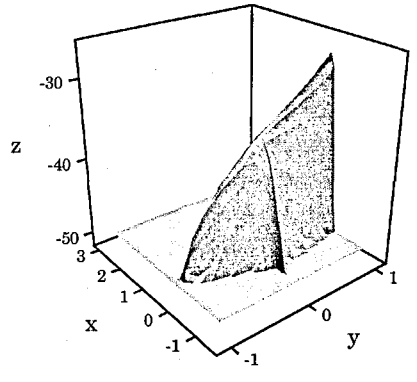

Figure 1.9 Constrained log-likelihood surface under the centred parametrization, truncated at $z = -50$, for the studentized values of the frontier data: $x = \mu_s$, $y = \gamma_1$, z = value of (1.4.14).

The boundary estimate of γ_1 is inevitable given the form of the reparametrization and the previously identified boundary estimate for δ. The ML estimates of the other two parameters are $\hat{\mu}$ = 0.8860 and $\hat{\sigma}$ = 0.7474. Contrasting these results with the MM estimates $\tilde{\mu}$ = 0.8849, $\tilde{\sigma}$ = 0.7488 and $\tilde{\gamma}_1$ = 0.9022, we see that there is little difference between the estimates of the location

and scale parameters obtained via the two methods. However, the method of moments estimate $\tilde{\gamma}_1 = 0.9022$ is numerically much closer to the true underlying value of $\gamma_1 = 0.8510$. This estimate for γ_1 corresponds to an estimate for λ of approximately 6.4, whilst the boundary maximum likelihood estimate results in an infinite estimate for λ. On the one hand, these wildly disparate numerical values are somewhat misleading, as a consideration of Figure 1.1 attests. In fact, the differences between the probability density functions corresponding to the three λ-values concerned are not that great. Nevertheless, the ML solution implies the existence of a hard lower threshold, whereas that for MM estimation does not.

Generally, in situations where a boundary estimate arises for γ_1 then, depending on whether γ_1 is of interest, we might carry out ML based inference for a general half-normal distribution or apply computer intensive methods as in Section 1.4.1.2. We see the re-estimation approach to dealing with boundary estimates of Azzalini & Capitanio (1999) as one based on an interpretation of the likelihood which is difficult to defend on objective grounds.

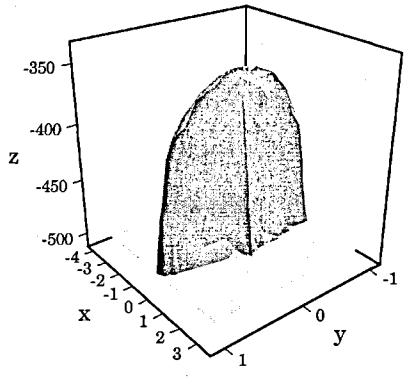

Figure 1.10 Constrained log-likelihood surface under the centred parametrization, truncated at $z = -500$, for the studentized values of the same simulated sample of size 500 from the $N(0,1) \equiv SN_D(0,1,0) \equiv SN_C(0,1,0)$ distribution used in Figure 1.6: $x = \mu_s$, $y = \gamma_1$, z = value of (1.4.14).

In Figure 1.10 we plot the surface defined by (1.4.14) for the simulated data from the standard normal distribution used previously to produce Figure 1.6. It can be seen that the reparametrization has removed the problematic plateau feature evident in Figure 1.6, attributed to the parameter redundancy of the direct parametrization. The surface has a unique global maximum and its contours

roughly approximate tight ellipses in shape. In both Figures 1.9 and 1.10 the ridges lie closely around the $\mu_s = 0$ axis. This value of μ_s corresponds to a value of $\mu = \bar{y}$ for the non-studentized data, which is the same as the MM estimate of μ. We might therefore expect the MM estimate of μ_s, $\tilde{\mu}_s = 0$, to provide a good initial estimate for the optimization of (1.4.14).

The main observations drawn from Figures 1.9 and 1.10 partly confirm points made by Azzalini (1985), Azzalini & Dalla Valle (1996) and Azzalini & Capitanio (1999). In the latter, a summary is given of the improvements brought about by the centred reparametrization, expressed mainly in terms of profile log-likelihoods. These improvements apply equally well to the constrained likelihood (1.4.14), and can be summarized as follows. As there is no parameter redundancy under the centred parametrization, its parameters are estimable. The reparametrization leads to improved shape characteristics of the log-likelihood surface, which in turn means that convergence of appropriately chosen optimization methods to the global maximum is, in general, relatively swift.

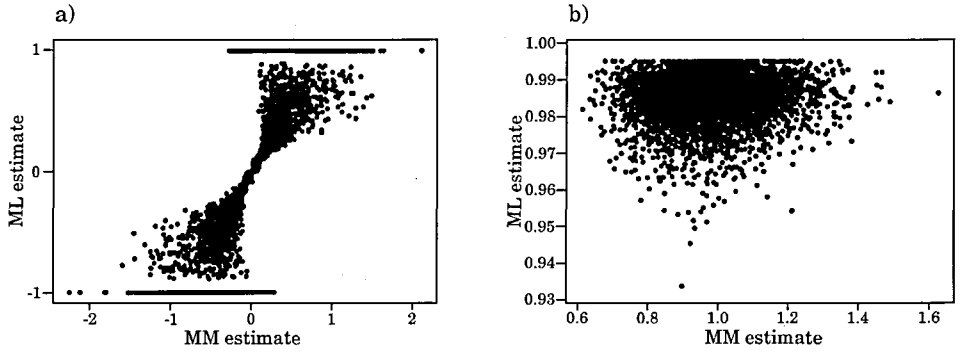

Figure 1.11 Scatterplots of the ML versus the MM estimate of γ_1 for 5000 simulated samples of size: a) 20 from the N(0,1) ≡ $SN_C(0,1,0)$ distribution; b) 500 from the $SN_D(0,1,20) \equiv SN_C(0.7969, 0.6041, 0.9851)$ distribution.

Azzalini & Capitanio (1999) gave a World Wide Web address from which two S-PLUS routines for fitting the skew-normal distribution using maximum likelihood estimation can be obtained. The routine sn.em uses the EM algorithm to maximize the log-likelihood whilst sn.mle employs gradient based methods. Both routines use the MM estimates as default starting values. As we have noted previously, $\tilde{\mu}$ usually provides a good initial estimate of $\tilde{\mu}$. However, in general, $\tilde{\gamma}_1$ and $\hat{\gamma}_1$ are

not strongly related, particularly if n is small or $|\gamma_1|$ is large. Another starting value needs to be used if $\tilde{\gamma}_1$ is inadmissible. Figure 1.11 illustrates the lack of any clear relation between $\hat{\gamma}_1$ and $\tilde{\gamma}_1$ using estimates obtained from simulated samples of size 20 from the N(0, 1) distribution and of size 500 from the $SN_D(0, 1, 20) \equiv SN_C(0.7969, 0.6041, 0.9851)$ distribution. The plotting symbols forming the horizontal lines in both scatterplots correspond to boundary ML estimates.

Experience shows that if the MM estimates are used as starting values then both routines can converge to a local, rather than the global, maximum of the log-likelihood. It has long been known that log-likelihood surfaces can contain multiple maxima, the problem being most acute for small samples. Schemes for identifying the global maximum in such circumstances date back to the early computational work of Barnett (1966). A standard approach is to use a grid of starting values in an attempt to ensure that the true global maximum is identified. An optional argument of the routine sn.mle allows the user to specify their own starting values and so can be used to carry out a grid search. The routine sn.em does not have this argument and so its use is strictly limited. Moreover, it can be very slow to execute and, more problematically, is based upon optimization for the direct parametrization which we identified in Section 1.4.1.2 as being parameter redundant for the normal case.

Our approach to finding ML estimates is based upon the optimization of (1.4.14) using the simplex algorithm of Nelder & Mead (1965). We use the starting value $\tilde{\mu}_s = 0$ and a grid of starting values for γ_1 spanning its full range. So as to ensure that the true global maximum has been identified, the maximum value found during optimization is contrasted with the values of (1.4.14) on the γ_1 boundary, namely $-\frac{1}{2}n\log_e\left(1+y_{s(1)}^2\right)$ and $-\frac{1}{2}n\log_e\left(1+y_{s(n)}^2\right)$, where $y_{s(1)}$ and $y_{s(n)}$ denote the minimum and maximum values of the studentized data, respectively.

A pilot simulation study incorporating the above approach showed that multiple maxima can occur on the constrained log-likelihood surface. Moreover, multiple maxima are most frequent for small samples from normal populations. The simulated sample in Table 1.2 provides a case in point. The log-likelihood associated with this sample has a local maximum 0.01, at ($\mu = 0.01$, $\sigma = 1.31$, $\gamma_1 = 0.10$) and a global maximum at ($\mu = -0.03$, $\sigma = 1.36$, $\gamma_1 = 0.73$). Using the default starting values, both sn.mle and sn.em converge to the local, rather than the global,

maximum. Figure 1.12 represents the shape of the constrained log-likelihood surface in the neighbourhood of these two maxima. Considering the problem in greater generality, approximately 5% of samples of size 20 from normal populations have log-likelihood surfaces which contain multiple maxima within the parameter space, whilst for samples from skew-normal populations with $\lambda = 20$ the frequency is close to 2%. For sample sizes as large as 500, multiple maxima are generally rare for all but those samples from close to normal populations. Even for this limiting case, their frequency of occurrence is only around 0.4%.

Table 1.2 Simulated sample of size 20 from the $SN_D(0,1,0) \equiv N(0,1)$ distribution.

2.583	1.343	−1.107	−1.203	−0.984	1.359	−0.799	1.580	−0.540	−0.918
1.121	0.412	−0.763	1.333	−1.928	0.873	−1.076	1.139	0.050	−2.258

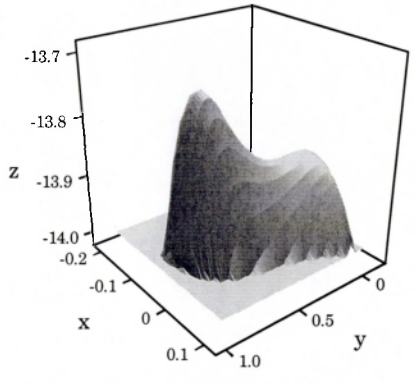

Figure 1.12 Constrained log-likelihood surface under the centred parametrization, truncated at $z = -14$, for the studentized values of the simulated sample of size 20 from the $SN_D(0,1,0) \equiv N(0,1)$ distribution given in Table 1.2: $x = \mu_s$, $y = \gamma_1$, $z =$ value of (1.4.14).

In Figures 1.13 and 1.14 we present the sampling distributions for the maximum likelihood estimates of the centred parameters for the same simulated data sets of size 20 and 500 used in the production of Figures 1.3, 1.4, 1.7 and 1.8. Comparisons of Figure 1.13 with Figure 1.7, and Figure 1.14 with Figure 1.8, prove most revealing. Firstly, there is very little appreciable difference between the sampling distributions of the method of moments and maximum likelihood

estimates of μ and σ. For small-sized samples and γ_1 close to 0, $\hat{\gamma}_1$ performs very poorly indeed. Figure 1.13c is dominated by spikes corresponding to boundary estimates. In comparison, the sampling distribution of $\tilde{\gamma}_1$ in Figure 1.7c is much more regular. Comparing Figure 1.14c with Figure 1.8c we see that the performance of $\hat{\gamma}_1$ approaches that of $\tilde{\gamma}_1$ as the sample size increases.

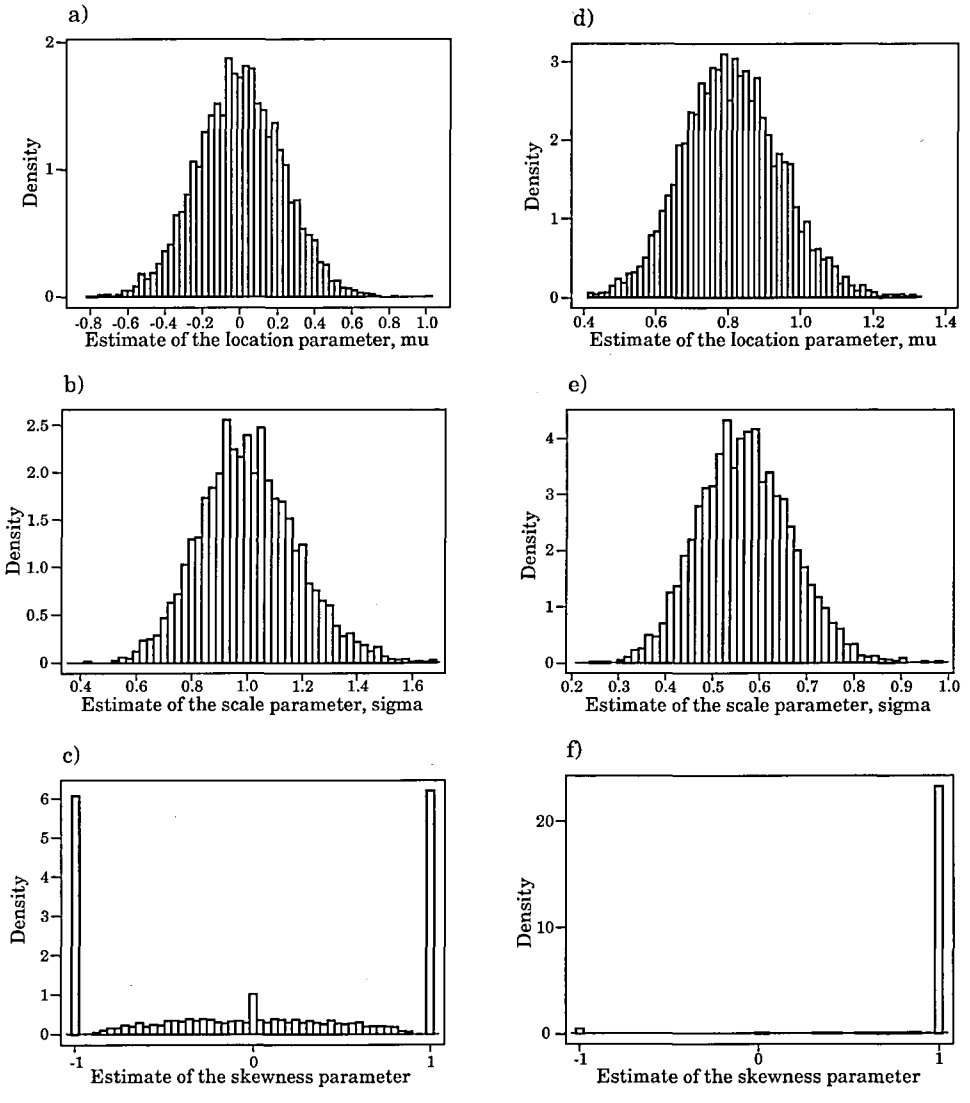

Figure 1.13 Empirical sampling distributions of the maximum likelihood estimates of μ, σ and γ_1 obtained from the same 5000 simulated samples of size 20 used to produce Figures 1.3 and 1.7: a), b), c) $SN_D(0,1,0) \equiv SN_C(0,1,0)$; d), e), f) $SN_D(0,1,20) \equiv SN_C(0.7969, 0.6041, 0.9851)$.

As can be seen from Figure 1.14f, the sampling distribution of $\hat{\gamma}_1$, like that of $\tilde{\gamma}_1$, is not normal for samples drawn from highly skewed populations, even for n as

large as 500. In passing, we note the spikes around 0 in the sampling distributions of $\hat{\gamma}_1$ displayed in Figures 1.13c and 1.14c. We attribute these to a compressing effect associated with the γ_1 scale, as on the λ scale they disappear. Whilst for large values of λ the sampling distribution of $\tilde{\gamma}_1$ is more regular (unimodal but, nevertheless, skewed), $\hat{\gamma}_1$ is superior in performance, even for small samples. Despite approximately 95% of the values of $\hat{\gamma}_1$ in Figure 1.13f being boundary estimates, the proportion of $\hat{\gamma}_1$ values in the neighbourhood of the true value of $\gamma_1 = 0.9851$ is far in excess of the corresponding proportion for $\tilde{\gamma}_1$.

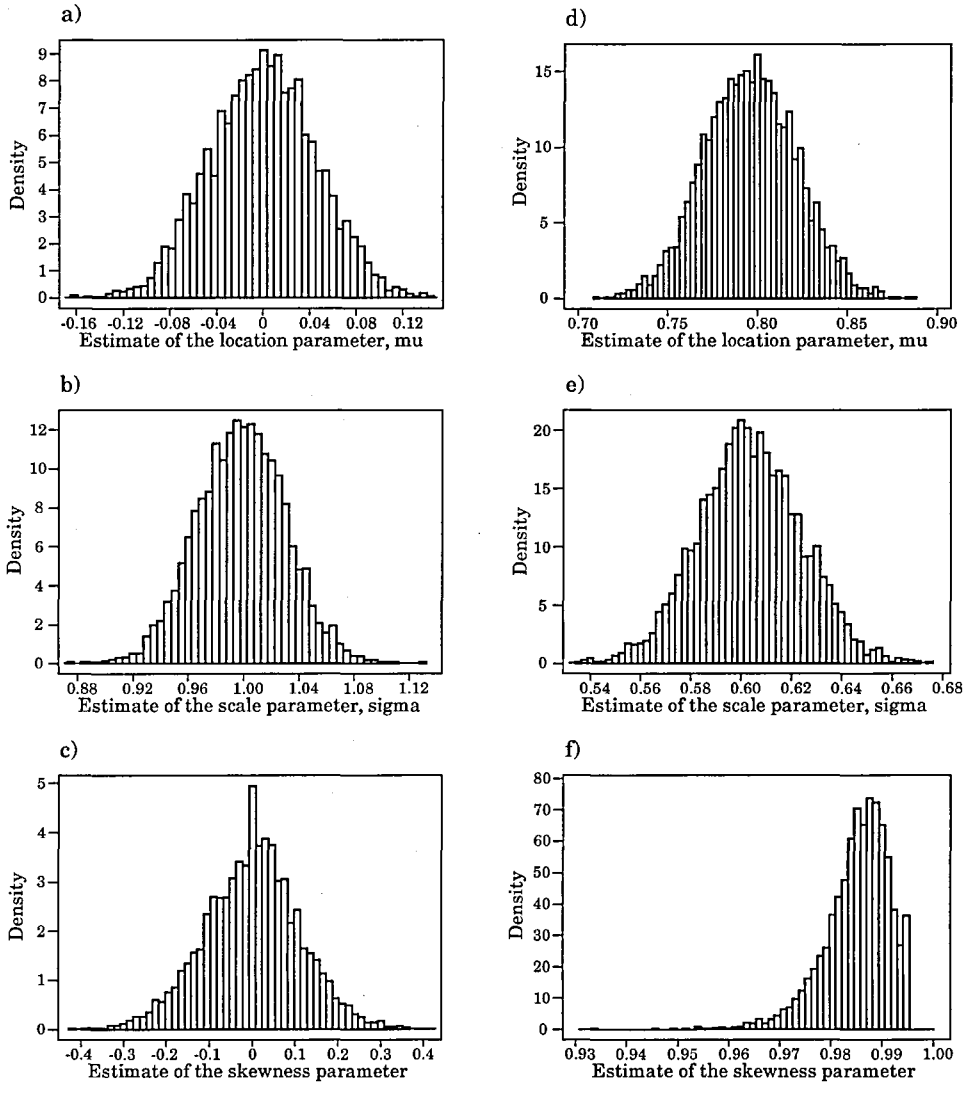

Figure 1.14 Empirical sampling distributions of the maximum likelihood estimates of μ, σ and γ_1 obtained from the same 5000 simulated samples of size 500 used to produce Figures 1.4 and 1.8: a), b), c) $SN_D(0,1,0) \equiv SN_C(0,1,0)$; d), e), f) $SN_D(0,1,20) \equiv SN_C(0.7969, 0.6041, 0.9851)$.

1.4.2.3 A Comparative Simulation Study

So as to compare the small-sample characteristics of the MM and ML estimators in greater detail, we conducted an in-depth simulation study. Samples of size n = 20, 50, 100, 200 and 500 were simulated from the $SN_D(0, 1, \lambda)$ distribution for λ = 0, 2, 5 and 20. For each (n, λ) combination, 5000 samples were simulated using the method of Henze (1986). As performance measures we used the mean value and mean squared error. For comparative purposes, the measures for $\tilde{\gamma}_1$ were calculated using truncation of inadmissible values to ± 0.99527. Clearly, the quoted mean squared errors for $\tilde{\gamma}_1$ are, in general, smaller than they would have been without such truncation. For the estimation of γ_1 we also recorded: the percentage of inadmissible MM estimates; the percentage of boundary ML estimates, and the percentage of samples for which the MM estimate was inadmissible and the ML estimate was a boundary estimate. The results obtained are presented in Tables 1.3-1.5, with each table representing the results for an individual parameter.

Table 1.3 Performance measures for the MM and ML estimates of μ from 5000 simulated samples of size n from the $SN_D(0,1,\lambda)$ distribution: mean; (mean squared error).

Sample size, n	$\lambda = 0$; $\mu = 0$		$\lambda = 2$; $\mu = 0.7136$		$\lambda = 5$; $\mu = 0.7824$		$\lambda = 20$; $\mu = 0.7969$	
	MM	ML	MM	ML	MM	ML	MM	ML
20	0.0022	0.0014	0.7157	0.7095	0.7826	0.7811	0.7984	0.8111
	(0.0480)	(0.0512)	(0.0248)	(0.0259)	(0.0201)	(0.0205)	(0.0185)	(0.0182)
50	0.0026	0.0029	0.7126	0.7077	0.7850	0.7780	0.7964	0.7967
	(0.0205)	(0.0210)	(0.0101)	(0.0104)	(0.0077)	(0.0077)	(0.0074)	(0.0070)
100	-0.0015	-0.0014	0.7146	0.7136	0.7818	0.7790	0.7971	0.7951
	(0.0099)	(0.0099)	(0.0049)	(0.0050)	(0.0039)	(0.0038)	(0.0038)	(0.0035)
200	0.0007	0.0007	0.7126	0.7123	0.7823	0.7818	0.7964	0.7943
	(0.0052)	(0.0052)	(0.0025)	(0.0025)	(0.0019)	(0.0019)	(0.0019)	(0.0017)
500	0.0000	0.0000	0.7146	0.7146	0.7819	0.7818	0.7963	0.7960
	(0.0020)	(0.0020)	(0.0010)	(0.0010)	(0.0007)	(0.0007)	(0.0007)	(0.0007)

From Table 1.3 we see there is little or no difference between the mean values and mean squared errors obtained for the method of moments and maximum likelihood estimates of μ.

Table 1.4 Performance measures for the MM and ML estimates of σ from 5000 simulated samples of size n from the $SN_D(0,1,\lambda)$ distribution: mean; (mean squared error).

Sample size, n	$\lambda = 0$; $\sigma = 1$		$\lambda = 2$; $\sigma = 0.7005$		$\lambda = 5$; $\sigma = 0.6228$		$\lambda = 20$; $\sigma = 0.6041$	
	MM	ML	MM	ML	MM	ML	MM	ML
20	0.9622	1.0058	0.6732	0.7009	0.5984	0.6083	0.5806	0.5691
	(0.0257)	(0.0307)	(0.0139)	(0.0160)	(0.0128)	(0.0120)	(0.0128)	(0.0107)
50	0.9871	0.9965	0.6887	0.6971	0.6142	0.6248	0.5929	0.5940
	(0.0101)	(0.0112)	(0.0057)	(0.0064)	(0.0052)	(0.0055)	(0.0054)	(0.0043)
100	0.9911	0.9922	0.6958	0.6973	0.6165	0.6206	0.5992	0.6027
	(0.0052)	(0.0052)	(0.0029)	(0.0030)	(0.0026)	(0.0027)	(0.0027)	(0.0022)
200	0.9965	0.9968	0.6970	0.6973	0.6204	0.6211	0.6008	0.6039
	(0.0025)	(0.0025)	(0.0014)	(0.0014)	(0.0013)	(0.0012)	(0.0013)	(0.0011)
500	0.9982	0.9982	0.6990	0.6991	0.6215	0.6217	0.6029	0.6035
	(0.0010)	(0.0010)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0004)

We know, of course, that $\tilde{\sigma}$ is a negatively biased estimator of σ. This fact is reflected in the results presented in Table 1.4. We note that the bias of $\hat{\sigma}$ is generally smaller than that of $\tilde{\sigma}$. For data from close to symmetric distributions, the mean squared error of $\tilde{\sigma}$ is generally smaller than that of $\hat{\sigma}$, while for highly skewed cases $\hat{\sigma}$ marginally outperforms $\tilde{\sigma}$ according to this criterion. Again, any differences between the mean values and mean squared errors of $\tilde{\sigma}$ and $\hat{\sigma}$ are very small.

From Table 1.5, $\hat{\gamma}_1$ generally outperforms $\tilde{\gamma}_1$, although the performance of $\hat{\gamma}_1$ is inferior for close to symmetric populations, particularly for small samples. The large mean squared error for these cases is consistent with the content of Figure 1.13c. In general, the frequencies of inadmissible and boundary estimates of γ_1 diminish with increasing n and as $\lambda \to 0$. However, boundary ML estimates are still possible for n as large as 500, albeit for samples from highly asymmetric populations. For such distributions the percentage of inadmissible method of moments estimates increases with sample size. The percentages appearing in the square brackets imply that there is little relation between the occurrence of inadmissible MM and boundary ML estimates of γ_1.

Table 1.5 Performance measures for the MM and ML estimates of γ_1 from 5000 simulated samples of size n from the $SN_D(0,1,\lambda)$ distribution: mean; (mean squared error); {percentage of inadmissible or boundary estimates, respectively}; [percentage of samples for which MM estimate was inadmissible and ML estimate was a boundary estimate].

Sample size, n	$\lambda = 0$; $\gamma_1 = 0$		$\lambda = 2$; $\gamma_1 = 0.4538$		$\lambda = 5$; $\gamma_1 = 0.8510$		$\lambda = 20$; $\gamma_1 = 0.9851$	
	MM	ML	MM	ML	MM	ML	MM	ML
20	-0.0015	0.0035	0.3171	0.4615	0.5806	0.7837	0.6769	0.9338
	(0.2064)	(0.5819)	(0.2081)	(0.4726)	(0.2045)	(0.2363)	(0.1947)	(0.0905)
	{3.74}	{49.16}	{8.86}	{54.74}	{21.70}	{77.24}	{29.40}	{95.08}
	[2.40]		[6.34]		[19.02]		[28.82]	
50	0.0014	-0.0003	0.3895	0.4464	0.7085	0.8448	0.7951	0.9811
	(0.1045)	(0.1812)	(0.1078)	(0.1351)	(0.0850)	(0.0405)	(0.0826)	(0.0049)
	{0.32}	{4.86}	{5.12}	{9.42}	{22.72}	{34.72}	{33.82}	{86.86}
	[0.12]		[1.14]		[10.02]		[30.26]	
100	-0.0002	-0.0012	0.4258	0.4467	0.7674	0.8474	0.8583	0.9851
	(0.0560)	(0.0731)	(0.0643)	(0.0576)	(0.0461)	(0.0138)	(0.0406)	(0.0007)
	{0.02}	{0.06}	{2.34}	{0.78}	{22.58}	{8.80}	{37.74}	{65.70}
	[0]		[0.02]		[2.42]		[25.34]	
200	-0.0013	-0.0015	0.4391	0.4473	0.8072	0.8499	0.8954	0.9857
	(0.0297)	(0.0339)	(0.0344)	(0.0266)	(0.0246)	(0.0054)	(0.0219)	(0.0002)
	{0}	{0}	{0.40}	{0}	{18.10}	{0.54}	{38.92}	{32.78}
	[0]		[0]		[0.22]		[13.08]	
500	-0.0002	-0.0004	0.4460	0.4505	0.8365	0.8494	0.9347	0.9852
	(0.0118)	(0.0122)	(0.0144)	(0.0104)	(0.0120)	(0.0019)	(0.0083)	(0.0000)
	{0}	{0}	{0.02}	{0}	{11.28}	{0}	{41.68}	{3.30}
	[0]		[0]		[0]		[1.26]	

In many practical situations, interest will focus on μ and σ, with γ_1 being a nuisance parameter. Our results indicate that for these situations there is little or no benefit in using ML estimation, and the extreme simplicity of MM estimation strongly favours its adoption. Should a complete specification of the underlying distribution be required then ML estimation is generally preferable. However, MM estimation performs better for small samples from close to symmetric populations.

1.4.3 Tests for Limiting Cases

The normal, half-normal and negative half-normal distributions warrant special attention as they can be specified in terms of two parameters rather than the three of the skew-normal class. Parsimony dictates that, for data displaying a high degree of asymmetry, or symmetry, we should investigate the appropriateness of the relevant limiting case. Salvan (1986), has shown that g_1 is the locally most

powerful location and scale invariant statistic for testing for normality within the skew-normal class. Here we consider significance tests for departures from an underlying half-normal distribution. Those for its negative analogue are then obvious.

In principal, one could define a generalized likelihood ratio based procedure to test for an underlying half-normal distribution against the alternative of some other member of the skew-normal class. In terms of the parameter γ_1, a test of these two hypotheses is equivalent to testing $H_0 : \gamma_1 = 0.99527$ against $H_1 : \gamma_1 < 0.99527$. However, as the value of γ_1 under the null hypothesis is a boundary point of the parameter space, the asymptotic distribution of the deviance for such a test is not χ_1^2 as given by standard likelihood theory. One could seek an empirical approximation to the sampling distribution of the deviation using simulation, as in Brooks *et al.* (1997), but we do not pursue that option here.

Instead, we consider a simple test based on the sample coefficient of skewness $g_1 = \tilde{\gamma}_1$. From Theorem 1.1, the asymptotic distribution of g_1 for data from a half-normal distribution is normal with mean 0.99527 and variance $8.03572n^{-1}$. A large-sample test follows immediately. However, this large-sample test should be used with caution because, as we have noted in Section 1.4.2.1, the sampling distribution of g_1 is not well approximated by the normal distribution, even for very large samples. In the absence of a better approximation to the sampling distribution of g_1, we propose a computer intensive alternative. As Barnard (1963) explains, a Monte Carlo based approach to significance testing is always available so long as data from the null model can be simulated. Here, such a test can be based on the rank of g_1 for the original data when ordered amongst the values of g_1 for samples of the same size simulated from the standard half-normal distribution.

Considering once more the percentages in square brackets in Table 1.5, and the content of Figure 1.11b, we note the following. If we use g_1 for testing for departures from a half-normal distribution, and base subsequent estimation upon the ML criterion, it is possible that the test might reject the null hypothesis and yet ML estimation could lead to the contradictory result of a boundary estimate for γ_1. Faced with this situation, the data analyst is advised to estimate the parameters

using both methods and compare the resultant densities graphically and for goodness-of-fit in order to check both for gross disparities between the two and for any lack of fit. On purely physical grounds, should we have reason to think that there is a threshold value associated with the variable of interest then this would tend to favour the adoption of the ML estimates. If the contrary were true, the MM solution would appear the more appropriate.

1.4.4 Two Illustrative Examples

In this section, we consider a number of issues associated with fitting the skew-normal distribution using two real data sets drawn from the statistical literature.

1.4.4.1 Glass Fibre Strength Data

In our first illustrative example we analyze a sample of experimental data introduced by Smith & Naylor (1987) and reproduced in Table 1.6. The data consist of 63 measurements on the strength of 1.5 cm lengths of glass fibre made at the National Physical Laboratory in England. Smith & Naylor consider the three-parameter Weibull distribution as a potential model for these data.

Table 1.6 The glass fibre strength data of Smith & Naylor (1987) ordered from smallest to largest.

0.55	0.74	0.77	0.81	0.84	0.93	1.04	1.11	1.13	1.24	1.25	1.27	1.28	1.29
1.30	1.36	1.39	1.42	1.48	1.48	1.49	1.49	1.50	1.50	1.51	1.52	1.53	1.54
1.55	1.55	1.58	1.59	1.60	1.61	1.61	1.61	1.61	1.62	1.62	1.63	1.64	1.66
1.66	1.66	1.67	1.68	1.68	1.69	1.70	1.70	1.73	1.76	1.76	1.77	1.78	1.81
1.82	1.84	1.84	1.89	2.00	2.01	2.24							

The method of moments and maximum likelihood estimates for the two parametrizations are:

$$\tilde{\xi} = 1.92,\ \hat{\xi} = 1.85;\ \tilde{\eta} = 0.52,\ \hat{\eta} = 0.47;\ \tilde{\lambda} = -6.30,\ \hat{\lambda} = -2.68,$$

and

$$\tilde{\mu} = 1.51,\ \hat{\mu} = 1.50;\ \tilde{\sigma} = 0.32,\ \hat{\sigma} = 0.31;\ \tilde{\gamma}_1 = -0.90,\ \hat{\gamma}_1 = -0.62,$$

respectively, the estimates of the direct parameters having been calculated from those of the centred ones using (1.4.5). Although the estimates of the location and scale parameters do not differ greatly under either parametrization, the differences between the estimates of the skewness parameters are relatively large. This reflects our earlier findings in Section 1.4.2.2 where the lack of any clear

relationship between the method of moments and maximum likelihood estimates of these parameters was discussed. In Figure 1.15 the densities corresponding to the method of moments and maximum likelihood estimates are superimposed on a histogram of the data. Whilst there is little difference between the lower tails of the two densities, the density fitted via maximum likelihood estimation appears to provide a more reasonable estimate of the mode and follow more closely the sample distribution in its upper tail. The visual impression that both densities under-represent the degree of kurtosis evident in the sample distribution, is partly confirmed by the chi-squared goodness-of-fit test. The chi-squared statistics for the MM and ML solutions, based on 10 class intervals with expected frequencies in excess of 5, corresponded to p-values of 0.07 and 0.11, respectively. Thus, using this criterion the ML solution is marginally superior. For both solutions, the major contributions to the chi-squared statistic were associated with the disparity between the fitted densities and the observed frequencies in the neighbourhood of the mode.

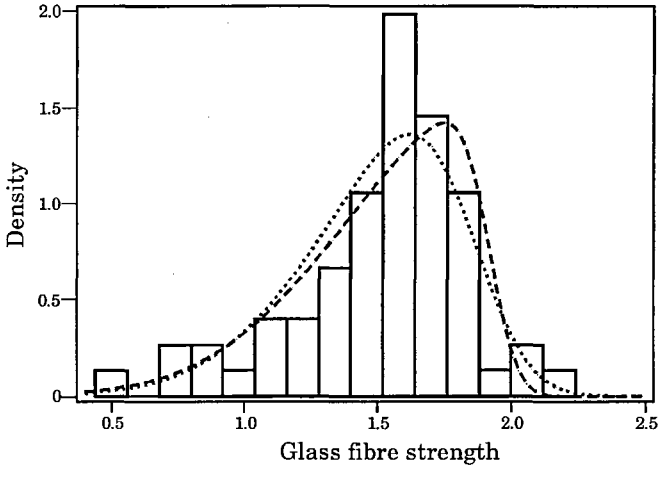

Figure 1.15 Histogram of the glass fibre strength data with superimposed skew-nomal densities fitted using the method of moments (– – –) and the method of maximum likelihood (·····).

1.4.4.2 Body Fat Measurements of Elite Athletes

Often, the skewness of a sample distribution is a consequence of mixing data from two or more distinct sub-populations. If this is the case, then a generally more informative analysis results from modelling the variable of interest within the various sub-samples rather than treating the data as a sample drawn from a single population. Our second data set, taken from Cook & Weisberg (1994), provides a case in point. The analysis we present also illustrates an issue raised towards the

end of Section 1.4.3. The data, reproduced in Table 1.7, consist of percentage body fat measurements made on 102 male and 100 female elite athletes representing ten different sports with highly disparate physiological demands. The athletes concerned trained at the Australian Institute of Sport.

Table 1.7 The percentage body fat measurements of 202 Australian Institute of Sport athletes, ordered from smallest to largest. The underlined values are those for the 102 male athletes.

5.63	5.80	5.90	5.93	6.00	6.00	6.03	6.06	6.06	6.10	6.16	6.20	6.26
6.33	6.33	6.43	6.46	6.53	6.56	6.56	6.59	6.76	6.82	6.86	6.92	6.96
6.99	7.06	7.16	7.19	7.19	7.22	7.29	7.35	7.35	7.42	7.49	7.52	7.68
7.72	7.82	7.88	8.07	8.18	8.44	8.45	8.47	8.51	8.51	8.51	8.54	8.56
8.61	8.64	8.77	8.84	8.84	8.87	8.87	8.94	8.97	8.97	9.00	9.02	9.03
9.10	9.17	9.20	9.20	9.36	9.40	9.40	9.50	9.53	9.56	9.56	9.56	9.79
9.86	9.89	9.91	10.05	10.05	10.12	10.15	10.16	10.25	10.48	10.53	10.64	10.74
10.81	11.05	11.07	11.07	11.22	11.29	11.47	11.50	11.63	11.64	11.66	11.72	11.77
11.79	11.85	11.95	12.16	12.20	12.39	12.55	12.61	12.78	12.92	13.06	13.35	13.46
13.49	13.61	13.91	13.93	13.97	14.26	14.52	14.53	14.69	14.98	15.01	15.07	15.31
15.58	15.59	15.95	16.20	16.25	16.38	16.58	16.86	17.07	17.22	17.24	17.41	17.51
17.64	17.71	17.71	17.89	17.93	17.95	18.04	18.08	18.08	18.14	18.48	18.72	18.77
19.17	19.20	19.26	19.35	19.39	19.51	19.61	19.63	19.64	19.75	19.83	19.88	19.94
19.99	20.10	20.12	20.43	20.86	21.30	21.30	21.32	21.47	21.79	22.25	22.39	22.43
22.62	23.01	23.01	23.11	23.30	23.66	23.70	23.70	24.69	24.88	24.97	25.16	25.26
26.50	26.57	26.65	26.78	28.83	30.10	35.52						

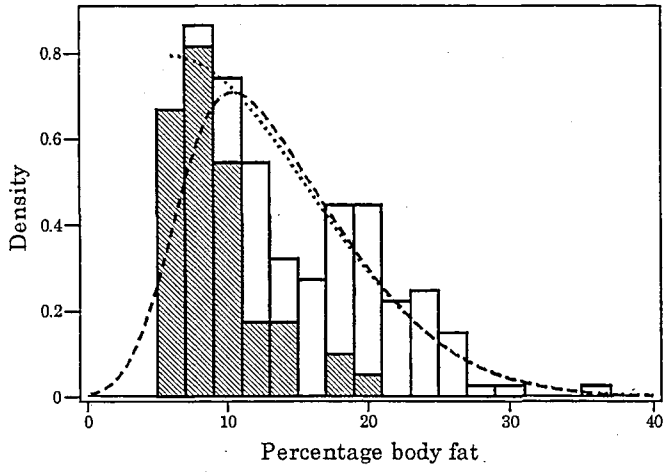

Figure 1.16 Histogram of the percentage body fat data with superimposed skew-normal densities fitted using the method of moments (– – –) and the method of maximum likelihood (· · · · ·). Shaded rectangles represent the density for male athletes; non-shaded rectangles, the density for females.

Figure 1.16 is a histogram of the data in which shading has been used to represent the density of male athletes in each class interval. The superimposed densities are those obtained from fitting the skew-normal distribution to the data

treated as a single sample using MM and ML estimation. The corresponding estimates of the direct and centred parameters are:

$$\tilde{\xi} = 6.04,\ \hat{\xi} = 5.63;\ \tilde{\eta} = 9.69,\ \hat{\eta} = 10.00;\ \tilde{\lambda} = 3.73,\ \hat{\lambda} = +\infty,$$

and

$$\tilde{\mu} = 13.51,\ \hat{\mu} = 13.62;\ \tilde{\sigma} = 6.17,\ \hat{\sigma} = 6.03;\ \tilde{\gamma}_1 = 0.76,\ \hat{\gamma}_1 = 0.99527.$$

We note the vast difference, both numerically and interpretationally, between $\tilde{\lambda}$ and $\hat{\lambda}$. The value for $\tilde{\lambda}$ of 3.73 would not lead us to suspect that a half-normal distribution was the parent population, whereas the message from $\hat{\lambda}$ is clearly that it is. When we applied the Monte Carlo significance test of Section 1.4.3 with 4999 simulated samples we found that the value 0.76 corresponded to the 13th percentile of the sampling distribution of $\tilde{\gamma}_1$. Thus, according to the observed value of $\tilde{\gamma}_1$, a half-normal distribution is a possible, though unlikely, model for the data.

Whilst the forms taken by the two fitted densities in Figure 1.16 are very similar across approximately 85% of the range of the data, the differences between their lower tails are important. It is known from the sports physiology literature that elite athletes seldom have less than 5% body fat. Generally, a minimum of between 3 and 4% body fat is necessary in order merely to survive. As can be seen from Table 1.7, the minimum percentage body fat measurement within the sample is 5.63, which is the threshold value fitted under ML estimation. Thus, the MM solution ascribes non-zero probability to physically unattainable measurements whilst the ML solution appears to over-estimate the threshold. Of course, we could have included the background information in the estimation process and fitted a half-normal distribution to the data with the value of ξ constrained to be some hard threshold value of, say, 3%.

On purely objective statistical grounds, neither solution, in fact, provides an adequate fit to the data. In addition to the observations made above, we note from Figure 1.16 that the sample distribution appears to have more than one mode. The chi-squared goodness-of-fit statistic for the MM solution, based on 13 class intervals with expected frequencies in excess of 5, corresponded to a p-value of zero to four decimal places. The major contributions to the lack-of-fit were found to be associated with the disparity between the fitted density and the observed frequencies in the lower tail area and in the neighbourhood of the smaller mode formed by body fat measurements of around 20%. The p-value for the equivalent

test for the ML solution was 0.01, again highly significant. For this solution, the major contributions to the test statistic corresponded to the disparities between the observed and expected frequencies in the neighbourhoods of the major mode and the two minor modes formed by body fat measurements of around 20% and 25%, respectively.

Whilst we have treated the data of this example as a single sample from a unimodal population, a primary reason for the poor fit of the skew-normal distribution is that the data do not actually arise from such a population. For instance, the values for the male and female athletes are markedly different, as can be appreciated from Figure 1.16. The generally higher percentage body fat measurements for the female athletes are in keeping with results from research in sports physiology; see, for example, Wilmore & Costill (1994, Chapter 16). Clearly, a major part of the skewness within the data results from mixing the measurements for the two sexes.

If the skew-normal distribution is fitted to the data for just the males we obtain the following MM and ML estimates for the two parametrizations:

$$\tilde{\mu} = 9.25,\ \hat{\mu} = 9.47;\ \tilde{\sigma} = 3.17,\ \hat{\sigma} = 2.90;\ \tilde{\gamma}_1 = 1.52786,\ \hat{\gamma}_1 = 0.99527.$$

and

$$\tilde{\xi} = 5.06,\ \hat{\xi} = 5.63;\ \tilde{\eta} = 5.26,\ \hat{\eta} = 4.81;\ \tilde{\lambda} = +\infty,\ \hat{\lambda} = +\infty,$$

where, in the calculation of the estimates of the direct parameters, the inadmissible MM estimate for γ_1 has been set equal to the maximum value of γ_1, namely 0.99527. Thus, both fits correspond to parent populations which are half-normal. The fit of the half-normal distribution to the data for the male athletes is considered in greater detail in Section 2.7 of Chapter 2.

When the skew-normal distribution is fitted to the measurements for the females, we obtain the following estimates:

$$\tilde{\mu} = 17.85,\ \hat{\mu} = 17.82;\ \tilde{\sigma} = 5.43,\ \hat{\sigma} = 6.74;\ \tilde{\gamma}_1 = 0.35,\ \hat{\gamma}_1 = 0.41.$$

and

$$\tilde{\xi} = 12.78,\ \hat{\xi} = 11.18;\ \tilde{\eta} = 7.42,\ \hat{\eta} = 9.46;\ \tilde{\lambda} = 1.65,\ \hat{\lambda} = 1.85.$$

In Figure 1.17, the densities corresponding to these estimates are superimposed upon a histogram of the measurements for the female athletes.

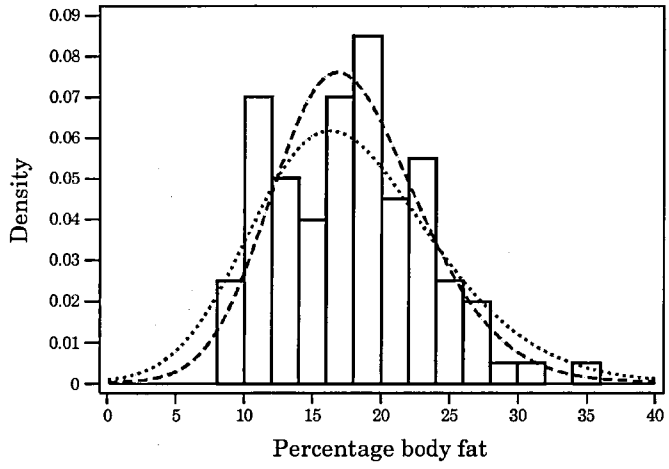

Figure 1.17 Histogram of the percentage body fat measurements of the 100 elite female athletes with superimposed skew-normal densities fitted using the method of moments (– – –) and the method of maximum likelihood (·····).

From a visual inspection of Figure 1.17 it would appear that the ML solution is far too heavy-tailed. The overall fit of the MM solution appears to be better, although the differences between it and the histogram perhaps indicate that the parent population is one with more than a single mode. These observations are supported by the results obtained from chi-squared goodness-of-fit analyses. The p-value of the chi-squared test for the MM solution, calculated using 10 class intervals with expected frequencies in excess of 5, formed from those used in the histogram, was 0.07. For this fit, the major contributions to the test statistic were those associated with disparities between the observed and expected frequencies around the first mode, and in the trough between modes, evident in the histogram. For an equivalent analysis based on the ML solution, the p-value of the test was 0.01, the major contributions to the goodness-of-fit statistic resulting from the disparities between the observed and expected frequencies in the tails and in the class intervals associated with the first and last mode of the histogram. We could proceed by splitting the data for the females further by sport type in an attempt to obtain a better fit and yet more insight into the factors influencing the body fat measurements of the athletes. However, we do not pursue that approach further here.

We conclude with the observation that, should we wish to fit a single density to all 202 measurements, we might contemplate an alternative approach based on finite mixture modelling. As the sample is clearly skew, the skew-normal

distribution suggests itself as a useful candidate model for the components used in such a mixture.

1.5 Summary and Directions for Future Research

In this last section of the chapter we summarize the main issues which have so far been addressed within in it, and indicate potential avenues for future research.

1.5.1 Summary

In Sections 1.2-1.3 the genesis, properties and univariate extensions of the general skew-normal distribution were reviewed.

The main section of the chapter, Section 1.4, dealt with some basic problems of inference associated with the general skew-normal distribution. Specifically, in Section 1.4.1, issues associated with estimating the distribution's direct parameters were addressed. In Section 1.4.1.1 we gave a simplified approach to calculating the MM estimates based on studentization of the original data, and highlighted unattractive features of the sampling distributions of the MM estimators. In this and subsequent sections, simulated data were used to illustrate the main issues raised. In Section 1.4.1.2 we explored problems associated with the shape of the log-likelihood function and traced a major shortcoming of the direct parametrization to its parameter redundancy for the important normal case.

Section 1.4.2 dealt with MM and ML based inference for the centred parametrization of the distribution. In Section 1.4.2.1 we derived the asymptotic distribution for the MM estimators and illustrated the improvement in the forms of their sampling distributions when compared with those of their direct parameter counterparts.

Competing numerical approaches for identifying the ML estimates were compared in Section 1.4.2.2, with particular attention being given to one employing a constrained version of the log-likelihood function and a grid based search utilizing the Nelder-Mead simplex. Some important deficiencies of two routines of Azzalini & Capitanio (1999) were discussed, these being related to the lack of any strong relation between the MM and ML estimates of γ_1, and the potential existence of multiple maxima on the log-likelihood surface. The proposed grid based search was found to resolve both of these shortcomings. We also illustrated improvements in the shape of the log-likelihood brought about by the reparametrization, some being consequences of its lack of parameter redundancy.

In Section 1.4.2.3 we presented the results of a simulation study designed to compare the small-sample characteristics of the MM and ML estimators of the centred parameters. We concluded that if γ_1 is a nuisance parameter, and we are interested in inference concerning μ and σ only, then MM estimation should be employed. If γ_1 is also of interest then ML estimation is, in general, preferable. However, MM estimation performs better for small samples from close to symmetric populations.

Procedures for testing for an underlying half-normal distribution were discussed in Section 1.4.3. We considered a large-sample test based on a normal approximation to the sampling distribution of the coefficient of skewness, and its Monte Carlo analogue.

Finally, in Section 1.4.6, two data sets were used to illustrate a number of issues associated with fitting the skew-normal distribution; the first example involving the strength of glass fibre, the second the body fat of elite athletes.

1.5.2 Directions for Future Research

In the foregoing sections of this chapter we have focused mainly on point estimation and tests for limiting cases of the general skew-normal distribution. Other forms of inference such as confidence set construction and hypothesis testing for pairs, or all three, of the distribution's parameters can be based on the distributional results for MM estimation presented in Theorem 1.1 and their ML counterparts given in Azzalini (1985). Standard asymptotic likelihood theory can be used to carry out inference for points within the parameter space. However, given the forms of the sampling distributions of the ML estimates, particularly that of $\hat{\gamma}_1$, one might envisage problems with the reliability of the results obtained. Clearly, standard asymptotic likelihood theory does not apply on the boundary of the parameter space. The generalized likelihood ratio procedure for testing for a half-normal parent population, referred to in Section 1.4.3, is another potential inferential tool which requires investigation. Taking a wider perspective, a comparison of the results obtained from MM and ML estimation with those arising from other methods of estimation would be of great interest.

As our examples in Section 1.4.4 show, the skew-normal class is somewhat limited in terms of its capacity to model kurtosis. In view of this shortcoming, the extended skew-normal class suggests itself as potentially being more relevant to the modelling of real data. From a consideration of the literature, the development

of inferential methods for this class of distributions appears to be a completely open field. The class's direct parametrization will suffer from the same problem of parameter redundancy as that identified for the general skew-normal distribution. Reparametrization in terms of $(\mu, \sigma, \gamma_1, \gamma_2)$ would appear to be an obvious potential remedy, but algorithms for actually locating the ML estimates of these parameters need to be developed. MM estimation is again trivial, and the results of Theorem 1.1 can be extended in the obvious way so as to cater for the additional parameter. Boundary problems will possibly be accentuated for this extended version of the skew-normal class, given that, in theory, boundary estimates can occur for both λ and ζ (or γ_1 and γ_2).

A further extension of the skew-normal class is to its multivariate counterpart studied by Azzalini & Dalla Valle (1996) and Azzalini & Capitanio (1999). The latter two authors have developed routines for fitting this multivariate skew-normal distribution which are available from the World Wide Web address `http://www.stat.unipd.it/dip/homes/azzalini/SN`. However, the parametrization used is an extension of the direct parametrization of the scalar class and it would appear that this parameterization could well suffer from parameter redundancy when the parent population is multivariate normal. Again, this potential problem needs to be investigated and an alternative parametrization identified should it be found to occur.

Returning to the univariate setting once more, the skew-normal class is just one of the potential skew classes that follow on applying Lemma 1.1. Mukhopadhyay & Vidakovic (1995) consider other classes with heavier tails, generated using the densities and distribution functions of the Laplace and t distributions. The (general) exponential and half-t distributions are the respective limiting distributions of these two classes when location and scale parameters are introduced. As will become clear, the treatment of the general half-normal distribution given in Chapter 2 is relevant to the development of inferential procedures for the parameters of these two limiting distributions. One could obviously define other classes of skew distributions using Lemma 1.1 in combination with the density and distribution function of any other symmetric distribution which might appeal, or by mixing the density of some chosen symmetric distribution with the distribution function of another. Thus, the further derivation of the wider class of possible distributions arising from Lemma 1.1 is a vast potential field for future research.

Lemma 1.1 is one means of generating skew distributions, but there are others. Jones & Faddy (2002) consider another general approach based on piecing together suitably scaled halves of symmetric component distributions, tracing the version of it involving normal distributions to Gibbons & Mylroie (1973). Another, less easily generalizable, option is to skew the density of a symmetric distribution directly by introducing appropriate additional parameters. This is the approach used by Jones (2001) to derive a skew extension of the t-distribution. As a general approach for producing skew multivariate distributions from spherically symmetric distributions, Jones (2002) proposes marginal replacement.

Together, these various techniques might be employed to derive a vast array of skew distributions in one or more dimensions. However, for the resulting distributions to be truly useful to the data analyst, sound inferential procedures need to be developed for the estimation and testing of their associated parameters.

Chapter 2 Large-sample Inference for the General Half-normal Distribution

2.1 Introduction

As we saw in Chapter 1, the general half-normal distribution is a limiting case of the general skew-normal class. Moreover, as we also saw, for highly skew data, both the MM and ML estimates of a fitted skew-normal distribution often correspond to a general half-normal parent population. We therefore consider the derivation of inferential results for the general half-normal distribution as being of genuine relevance to the modelling of skew data. Surprisingly, nothing had been published regarding inference for the general half-normal distribution prior to the appearance of Pewsey (2002a). In the sequel, the presented material draws heavily on the content of that paper. However, we note that the important results presented here concerning bias-correction, and those for MM estimation, were not included in that publication.

The subsequent sections of the chapter are organized as follows. In Section 2.2 we consider the origins of the half-normal distribution and its extension to the general half-normal distribution. The numerous distributions for which the half-normal distribution is a special case are also identified. Section 2.3 addresses point estimation of the distribution's parameters. Specifically, we provide details of the estimates arising from the method of moments and the method of maximum likelihood. Large-sample estimation based on the resulting estimators is the theme of Section 2.4. Using the asymptotic results given in Theorem 1.1 of Chapter 1, in Section 2.4.1 we derive the asymptotic distribution of the MM estimators for the parameters of the general half-normal distribution. These results lead us to the definition of bias-corrected estimates and large-sample confidence sets for the distribution's parameters. Section 2.4.2 follows along similar lines, but deals with corresponding results for the ML based estimators. We first use extreme value theory to derive an asymptotic distributional result involving the ML estimator of the location parameter of the distribution. Subsequently, an asymptotic result for the distribution of the ML estimator of the distribution's scale parameter is

obtained. Bias-corrected estimators follow directly from these two results, as do constructions for large-sample confidence sets.

In Section 2.5 we present the details of a Monte Carlo experiment designed to investigate the sampling properties of the point estimators, and the coverages of the confidence sets, identified in Sections 2.3 and 2.4. The use of the confidence sets as a basis upon which to perform hypothesis testing is discussed briefly in Section 2.6. In Section 2.7 we illustrate the application of the developed methodology with an analysis of the body fat measurements of the male elite athletes introduced in Section 1.4.4.2 of Chapter 1.

The last section of the chapter, Section 2.8, provides a summary of the preceeding sections and the main conclusions drawn from them. The chapter ends with an indication of related potential lines for future research.

2.2 Daniel's Half-normal Distribution and its Extension

If Z is a standard normal random variable, $Z \sim \mathrm{N}(0,1)$, then $X = |Z|$ follows a (standard) half-normal distribution, and $-X = -|Z|$ a (standard) negative half-normal distribution. The standard half-normal distribution is a special case of the folded normal and truncated normal distributions (Johnson *et al.*, 1994, pp. 156, 170). It also arises as the central chi distribution with one degree of freedom (Johnson *et al.*, 1994, p. 417).

Extending the distribution via the inclusion of location and scale parameters, $Y = \xi + \eta X$ is a general half-normal random variable with density

$$\begin{aligned} f(y) &= \frac{2}{\eta}\phi\left(\frac{y-\xi}{\eta}\right) \\ &= \frac{b}{\eta}\exp\left\{\frac{-(y-\xi)^2}{2\eta^2}\right\}, \qquad y > \xi, -\infty < \xi < \infty, \ \eta > 0. \end{aligned} \tag{2.2.1}$$

where, as in Chapter 1, $b = (2/\pi)^{1/2}$ and $\phi(\cdot)$ denotes the standard normal density. We write $Y \sim \mathrm{HN}(\xi, \eta)$ to denote the fact. The variable $Y^* = \xi - \eta X$ follows a negative general half-normal distribution. Inference for random samples from such a distribution follows in an obvious way from that for random samples from a general half-normal distribution, as $-Y^* \sim \mathrm{HN}(-\xi, \eta)$.

The $\mathrm{HN}(0, \eta)$ distribution first appeared in Cuthbert Daniel's classic paper of 1959 introducing half-normal plots. Other relevant early work includes that of

Elandt (1961) who, considering the HN$(0,\eta)$ distribution as a special case of the folded normal class, gave the distribution's first four moments. The HN$(0,\eta)$ distribution has become a popular distributional model in two main contexts. In addition to providing a parent population from which to generate skew data sets in simulation studies, the distribution is also employed to describe the form of skew-distributed errors in stochastic frontier modelling; see, for example, Aigner *et al.* (1977). This particular case of the general half-normal distribution also arises as a special case of the generalized Rayleigh distribution (Johnson *et al.*, 1994, p. 453).

Finally, we note that, as well as being a limiting distribution of the skew-normal class, the HN(ξ,η) distribution is also a special case of both the generalized gamma distribution and the two parameter chi distribution (Johnson *et al.*, 1994, pp. 385, 454).

2.3 Point Estimation

In this section we consider the point estimation of the parameters of the general half-normal distribution using the method of moments and the method of maximum likelihood.

2.3.1 Method of Moments Estimation

The first two moments of $Y = \xi + \eta X$ are:

$$E(Y) = \xi + b\eta \quad \text{and} \quad E(Y^2) = \xi^2 + 2b\xi\eta + \eta^2. \tag{2.3.1}$$

Equating these two moments to their sample analogues for a random sample $\underline{y} = (y_1, ..., y_n)$ of n observations from the HN(ξ,η) distribution, one obtains the moment estimates

$$\tilde{\xi} = \bar{y} - \frac{bs}{\left(1-b^2\right)^{1/2}} \quad \text{and} \quad \tilde{\eta} = \frac{s}{\left(1-b^2\right)^{1/2}}, \tag{2.3.2}$$

where $\bar{y}$ and s denote the mean and standard deviation of $\underline{y}$, respectively. From (2.3.1), should the threshold parameter ξ be known then the moment estimate of η becomes $\tilde{\eta}_{\xi} = (\bar{y} - \xi)/b$. If, on the other hand, the scale parameter, η, is known then the estimate of ξ becomes $\tilde{\xi}_{\eta} = \bar{y} - b\eta$. Of these two possibilities, the former is the more likely to occur in practice. Inadmissible estimates are likely to occur

using either estimator of the location parameter ξ; an inadmissible estimate in this context being one which is greater than $y_{(1)}$, the smallest value in the sample $\underline{y}$.

2.3.2 Maximum Likelihood Estimation

From (2.2.1) the likelihood function for (ξ,η), given $\underline{y}$, is

$$L(\xi,\eta;\underline{y})=\left(\frac{b}{\eta}\right)^n \exp\left\{-\frac{1}{2\eta^2}\sum_{i=1}^{n}(y_i-\xi)^2\right\}\mathrm{hv}(y_{(1)}-\xi), \qquad (2.3.3)$$

where $\mathrm{hv}(\cdot)$ is the unit Heaviside function. In order to maximize (2.3.3) we need to minimize $\sum_{i=1}^{n}(y_i-\xi)^2$ subject to the constraint $\xi \le y_{(1)}$. This obviously occurs when $\hat{\xi}=y_{(1)}$. For this choice of ξ the log-likelihood function is maximized when $\hat{\eta}=\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i-y_{(1)})^2\right\}^{1/2}$. Should ξ be known, then the ML estimate of η is $\hat{\eta}_\xi=\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i-\xi)^2\right\}^{1/2}$, whilst $\hat{\xi}=y_{(1)}$ is also the ML estimate $\hat{\xi}_\eta$ of ξ for the case where η is known.

2.4 Asymptotic Distributions, Bias-correction and Large-sample Confidence Sets

In this section we obtain asymptotic distributional results for the MM and ML based estimators identified in Section 2.3. These results are then used to derive bias-corrected estimators and large-sample confidence sets for the parameters of the general half-normal distribution.

2.4.1 Moment Based Inference

From the distributional results given in Theorem 1.1 of Chapter 1, we have:

$$E(\bar{y})=\xi+b\eta,\quad \mathrm{var}(\bar{y})=\frac{\eta^2(1-b^2)}{n},$$

$$E(s)=\eta(1-b^2)^{1/2}\left\{1-\frac{(3+\beta_2)}{8n}\right\}+O(n^{-3/2}),$$

$$\mathrm{var}(s)=\frac{\eta^2(1-b^2)(\beta_2-1)}{4n}+O(n^{-3/2}),$$

$$\operatorname{cov}(\bar{y},s)=\frac{\eta^2(1-b^2)\gamma_1}{2n}+O(n^{-3/2}),$$

where γ_1 = 0.9952717 and β_2 = 3.8691773. Using these results together with (2.3.2), it follows that:

$$\begin{aligned}
&E(\tilde{\xi})=\xi+\frac{c_1\eta}{n}+O(n^{-3/2}),\quad \operatorname{var}(\tilde{\xi})=\frac{c_2\eta^2}{n}+O(n^{-3/2}),\\
&E(\tilde{\eta})=\eta\left(1-\frac{c_3}{n}\right)+O(n^{-3/2}),\quad \operatorname{var}(\tilde{\eta})=\frac{c_4\eta^2}{n}+O(n^{-3/2}),\\
&\operatorname{cov}(\tilde{\xi},\tilde{\eta})=\frac{c_5\eta^2}{n}+O(n^{-3/2}),
\end{aligned}\tag{2.4.1}$$

where c_1 = 0.6851013, c_2 = 0.3413251, c_3 = 0.8586472, c_4 = 0.7172943 and c_5 = −0.2723381. Appealing to Lemma 1.2, the joint distribution of $(\tilde{\xi},\tilde{\eta})$ is asymptotically bivariate normal.

The results for $E(\tilde{\eta})$ and $E(\tilde{\xi})$ in (2.4.1) suggest we might contemplate the use of the following bias-corrected point estimates for η and ξ,

$$\tilde{\eta}_{BC}=\frac{s}{(1-b^2)^{1/2}\left(1-\frac{c_3}{n}\right)}\quad\text{and}\quad \tilde{\xi}_{BC}=\bar{y}-\frac{s}{(1-b^2)^{1/2}}\left\{b+\frac{c_1}{(n-c_3)}\right\}.\tag{2.4.2}$$

Using the results for $\tilde{\xi}$, and without including any bias-correction in the estimation of η, the limits of a conventional approximate $100(1-\alpha)\%$ confidence interval for ξ are given by

$$\tilde{\xi}-\left\{\frac{c_1}{n}\pm z_{\alpha/2}\left(\frac{c_2}{n}\right)^{1/2}\right\}\tilde{\eta},\tag{2.4.3}$$

where $z_{\alpha/2}$ denotes the upper $\alpha/2$ quantile of the standard normal distribution. Allowing for bias-correction in the estimation of η, the limits of the equivalent interval for ξ are

$$\tilde{\xi}-\left\{\frac{c_1}{n}\pm z_{\alpha/2}\left(\frac{c_2}{n}\right)^{1/2}\right\}\tilde{\eta}_{BC}.\tag{2.4.4}$$

From the results for $\tilde{\eta}$, the limits of a conventional approximate $100(1-\alpha)\%$ confidence interval for η are given by

$$\frac{\tilde{\eta}}{1-\frac{c_3}{n}\pm z_{\alpha/2}\left(\frac{c_4}{n}\right)^{1/2}}. \tag{2.4.5}$$

Making use of the results for the asymptotic joint distribution of $(\tilde{\xi},\tilde{\eta})$, a large-sample approximate $100(1-\alpha)\%$ simultaneous confidence region for (ξ,η) is given by $\left\{(\xi,\eta): U < \chi^2_{2,1-\alpha}\right\}$ where

$$U=\frac{n}{\eta^2}\left[c_6\left(\tilde{\xi}-\xi-\frac{c_1\eta}{n}\right)^2+2c_7\left(\tilde{\xi}-\xi-\frac{c_1\eta}{n}\right)\left\{\tilde{\eta}-\eta\left(1-\frac{c_3}{n}\right)\right\}+2\left\{\tilde{\eta}-\eta\left(1-\frac{c_3}{n}\right)\right\}^2\right], \tag{2.4.6}$$

c_6 = 4.2029975, c_7 = 1.5957691 and $\chi^2_{2,1-\alpha}$ is the $(1-\alpha)$ quantile of the chi-squared distribution with two degrees of freedom.

The asymptotic distributions of $\tilde{\xi}_\eta$ and $\tilde{\eta}_\xi$, obtained on applying the central limit theorem, are

$$\mathrm{N}\left(\xi,\frac{(1-b^2)\eta^2}{n}\right) \quad \text{and} \quad \mathrm{N}\left(\eta,\frac{(1-b^2)\eta^2}{b^2 n}\right),$$

respectively. The construction of individual large-sample confidence intervals for ξ (when η is known) and η (when ξ is known) is then trivial.

2.4.2 Likelihood Based Inference

Likelihood based inference for the general half-normal distribution is problematic in that the standard regularity conditions underlying it cannot be appealed to. Specifically, as is evident from (2.2.1), the support of the density depends on the location parameter ξ. General issues of likelihood based inference for irregular problems are reviewed by Smith (1985, 1989) and Cheng & Traylor (1995). A standard approach often used in an attempt to circumvent the difficulties associated with the type of irregularity experienced here, is to reparametrize the distribution. For instance, in the present case one might contemplate employing a parametrization equivalent to that of the centred parametrization for the skew-normal distribution, that is,

$$Y = \mu + \sigma\left[\frac{X - E(X)}{\{\text{var}(X)\}^{1/2}}\right], \quad -\infty < \mu < \infty, \ \sigma > 0,$$

leading to the definition of a general half-normal distribution with mean μ and variance σ^2. However, as Smith & Naylor (1987) observe, such a reparametrization is of no practical benefit if the original parameters are of paramount interest. This is indeed the case for the general half-normal distribution, as the problematic threshold value, ξ, will usually be of particular practical interest.

Given that standard likelihood theory cannot be used, we consider an alternative approach to obtaining distributional results for the ML estimators. First we use extreme value theory to derive an asymptotic result involving the ML estimator of ξ. For $Y \sim \text{HN}(\xi, \eta)$,

$$F_Y(y) = \begin{cases} 2\Phi\left(\dfrac{y-\xi}{\eta}\right) - 1 & \xi \le y < \infty \\ 0 & y < \xi \end{cases},$$

and thus the distribution function of $\hat{\xi} = Y_{(1)}$ is given by

$$\begin{aligned} F_{\hat{\xi}}(y) &= 1 - \{1 - F_Y(y)\}^n \\ &= \begin{cases} 1 - 2^n\left\{1 - \Phi\left(\dfrac{y-\xi}{\eta}\right)\right\}^n & \xi \le y < \infty \\ 0 & y < \xi \end{cases}, \end{aligned}$$

which has a degenerate limiting distribution at $y = \xi$ as $n \to \infty$. Using standard extreme value theory (see, for example, Arnold *et al.* (1992)), as $Y_{(1)}$ is limited on the left, its extreme value distribution is of the Weibull type. Given $F_Y(y)$, the $1/n$ quantile, or so-called smallest characteristic value of Y, s_n, is obtained as $s_n = \eta\,\Phi^{-1}\left(\frac{1}{2} + \frac{1}{2n}\right) + \xi$. Hence, we consider the limiting distribution of

$$W_n = \frac{\hat{\xi} - \xi}{\left\{\eta\,\Phi^{-1}\left(\frac{1}{2} + \frac{1}{2n}\right) + \xi\right\} - \xi} = \frac{\hat{\xi} - \xi}{\eta\,\Phi^{-1}\left(\frac{1}{2} + \frac{1}{2n}\right)}.$$

Now,

$$\lim_{y\to 0}\frac{F_Y(ky+\xi)}{F_Y(y+\xi)}=\lim_{y\to 0}\frac{2\Phi\left(\frac{ky}{\eta}\right)-1}{2\Phi\left(\frac{y}{\eta}\right)-1}$$

$$=\lim_{y\to 0}\frac{2\frac{k}{\eta}\phi\left(\frac{ky}{\eta}\right)}{2\frac{1}{\eta}\phi\left(\frac{y}{\eta}\right)}=k,$$

the last expression obtained using l'Hôpital's rule. Thus the extreme value distribution of W_n is Weibull with parameter $\gamma=1$, i.e. it is exponential with parameter $\lambda=1$. Summarizing,

$$\frac{\hat{\xi}-\xi}{\eta\,\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right)}\sim \text{EXP}(1). \tag{2.4.7}$$

Table 2.1 Values of $\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right)$ and $1/(bn)$ for a range of sample sizes.

Sample size, n	$\left(\frac{1}{2}+\frac{1}{2n}\right)$	$\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right)$	$1/(bn)$
20	0.5250	0.0627068	0.0626657
30	0.5167	0.0417893	0.0417771
40	0.5125	0.0313380	0.0313329
50	0.5100	0.0250689	0.0250663
100	0.5050	0.0125335	0.0125331

We can obtain an approximation to $z=\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right)$ as follows. Using a series expansion of $\phi(x)$ about 0,

$$\frac{1}{2n}=\frac{1}{(2\pi)^{1/2}}\int_0^z \exp\left(-x^2/2\right)dx$$

$$=\frac{1}{(2\pi)^{1/2}}\int_0^z\left(1-\frac{x^2}{2}+\frac{x^4}{8}-\ldots\right)dx$$

$$=\frac{1}{(2\pi)^{1/2}}\left[x-\frac{x^3}{6}+\frac{x^5}{40}-\ldots\right]_0^z,$$

the right hand side of which is approximately equal to $z/(2\pi)^{1/2}$ for z in the neighbourhood of 0. Thus, $z\approx(\pi/2)^{1/2}/n=1/(bn)$. This approximation is accurate to

2 decimal places for $n = 10$, and to 5 decimal places for $n = 50$. In Table 2.1 we give more precise values of $\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right)$, as well as the values of $1/(bn)$ to the same accuracy, for a pertinent range of sample sizes.

Turning to consider the distribution of $\hat{\eta}$, we have $n\hat{\eta}^2 = \sum_{i=1}^{n}\left(Y_i - Y_{(1)}\right)^2$ where $Y_i = \xi + \eta X_i$, $X_i = |Z_i|$ and $Z_i \sim \mathrm{N}(0,1)$. Hence, $n\hat{\eta}^2/\eta^2 = \sum_{i=1}^{n}\left(X_i - X_{(1)}\right)^2$, where $X_{(1)}$ is the minimum value of a random sample of size n, $X_1,\ldots,X_n$, from the standard half-normal distribution. As $n \to \infty$, $X_{(1)} \xrightarrow{P} 0$, and thus, asymptotically,

$$n\hat{\eta}^2/\eta^2 \sim \chi^2_{n-1}. \tag{2.4.8}$$

For finite sample sizes, $E\left(\hat{\eta}^2\right) \approx \frac{(n-1)}{n}\eta^2$, which suggests the use of the bias-corrected estimate,

$$\hat{\eta}_{BC} = \left(\frac{n}{n-1}\hat{\eta}^2\right)^{1/2} = \left\{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - y_{(1)}\right)^2\right\}^{1/2}. \tag{2.4.9}$$

Similarly, from (2.4.7) we have, for finite n, $E\left(\hat{\xi}\right) \approx \xi + \eta\,\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right)$, which leads us to the bias-corrected estimator

$$\hat{\xi}_{BC} = \hat{\xi} - \hat{\eta}_{BC}\,\Phi^{-1}\left(\tfrac{1}{2}+\tfrac{1}{2n}\right). \tag{2.4.10}$$

Using (2.4.7) without any bias-correction in the estimation of η, an approximate $100(1-\alpha)\%$ confidence interval for ξ is given by

$$y_{(1)} + \hat{\eta}\log_e\left(\tfrac{\alpha}{2}\right)\Phi^{-1}\left(\tfrac{1}{2}+\tfrac{1}{2n}\right) < \xi < y_{(1)} + \hat{\eta}\log_e\left(1-\tfrac{\alpha}{2}\right)\Phi^{-1}\left(\tfrac{1}{2}+\tfrac{1}{2n}\right). \tag{2.4.11}$$

Allowing for bias-correction in the estimation of η, the interval becomes

$$y_{(1)} + \hat{\eta}_{BC}\log_e\left(\tfrac{\alpha}{2}\right)\Phi^{-1}\left(\tfrac{1}{2}+\tfrac{1}{2n}\right) < \xi < y_{(1)} + \hat{\eta}_{BC}\log_e\left(1-\tfrac{\alpha}{2}\right)\Phi^{-1}\left(\tfrac{1}{2}+\tfrac{1}{2n}\right). \tag{2.4.12}$$

If η is known then a large-sample confidence interval for ξ with the same nominal confidence level is obtained by substituting η for $\hat{\eta}$ in (2.4.11).

From (2.4.8), an approximate $100(1-\alpha)\%$ confidence interval for η is given by

$$\left(\hat{\eta}\left\{\frac{n}{\chi^2_{n-1,1-\alpha/2}}\right\}^{1/2}, \hat{\eta}\left\{\frac{n}{\chi^2_{n-1,\alpha/2}}\right\}^{1/2}\right), \tag{2.4.13}$$

where $\chi^2_{n-1,\alpha/2}$ denotes the $\alpha/2$ quantile of the chi-squared distribution with $n-1$ degrees of freedom. When ξ is known, $n\hat{\eta}_\xi^2 = \sum_{i=1}^{n}(Y_i - \xi)^2 = \eta^2 \sum_{i=1}^{n} X_i^2$. Hence,

$$\frac{n\hat{\eta}_\xi^2}{\eta^2} \sim \chi^2_n, \tag{2.4.14}$$

and the construction of confidence intervals for η in this case follows in an obvious way.

From the derivation of the asymptotic distributional results summarized in (2.4.7) and (2.4.8), it follows that the distributions of the ML estimators $\hat{\xi}$ and $\hat{\eta}$ are asymptotically independent. In the construction of a large-sample confidence region for (ξ,η) we therefore assume the two pivotal statistics in (2.4.7) and (2.4.8) to be independent and proceed as in the classic construction of a Mood exact region for the parameters of a univariate normal distribution; see Mood (1950, p. 227) and Arnold & Shavelle (1998). Thus, for a conventional $100(1-\gamma)\%$ simultaneous confidence region we assume,

$$\begin{aligned}(1-\gamma) &= (1-\alpha)^2 \\ &\approx P\left\{-\log_e\left(1-\tfrac{\alpha}{2}\right) < \frac{\left(\hat{\xi}-\xi\right)}{\eta\,\Phi^{-1}\left(\frac{1}{2}+\frac{1}{2n}\right)} < -\log_e\left(\tfrac{\alpha}{2}\right),\ \chi^2_{n-1,\alpha/2} < \frac{n\hat{\eta}^2}{\eta^2} < \chi^2_{n-1,1-\alpha/2}\right\}.\end{aligned}$$

As a function of γ we can express α as $\alpha = 1-(1-\gamma)^{1/2}$. An approximate $100(1-\gamma)\%$ confidence region for (ξ,η) is then given by the set

$$\left\{(\xi,\eta): \hat{\eta}\left\{\frac{n}{\chi^2_{n-1,1-\alpha/2}}\right\}^{1/2} < \eta < \hat{\eta}\left\{\frac{n}{\chi^2_{n-1,\alpha/2}}\right\}^{1/2},\right.$$
$$\left.\hat{\xi} + \eta\log_e\left(\tfrac{\alpha}{2}\right)\Phi^{-1}\left(\tfrac{1}{2}+\tfrac{1}{2n}\right) < \xi < \hat{\xi} + \eta\log_e\left(1-\tfrac{\alpha}{2}\right)\Phi^{-1}\left(\tfrac{1}{2}+\tfrac{1}{2n}\right)\right\}. \tag{2.4.15}$$

To illustrate the forms taken by the confidence regions specified in (2.4.6) and (2.4.15), in Figure 2.1 we present the respective approximate 90% confidence regions for (ξ,η) corresponding to sample sizes of 20 and 100, calculated using $\tilde{\xi} = \hat{\xi} = 0$ and $\tilde{\eta} = \hat{\eta} = 1$.

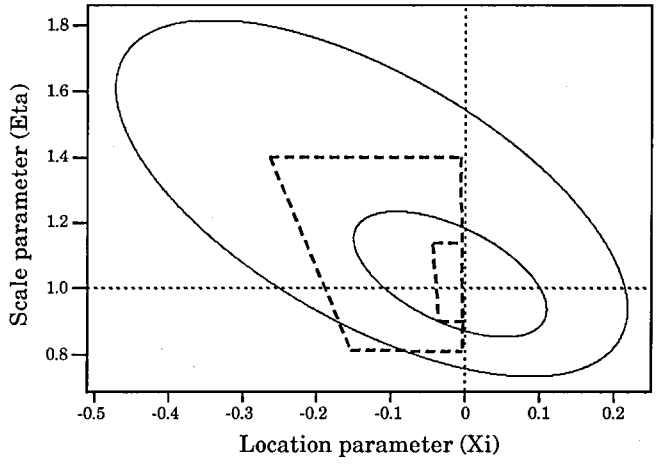

Figure 2.1 Approximate 90% confidence regions for (ξ,η) for sample sizes of 20 (large regions) and 100 (small regions) with $\tilde{\xi} = \hat{\xi} = 0$ and $\tilde{\eta} = \hat{\eta} = 1$. The ellipses are MM based regions calculated using (2.4.6); the rectilinear regions delimited by the heavy dashed lines are those from ML theory obtained using (2.4.15).

2.5 Monte Carlo Results

In order to investigate the small-sample characteristics of the point estimates and confidence sets identified in Sections 2.3 and 2.4, we conducted a simulation experiment. For a given sample size, n, ranging between 20 and 100, we generated one million pseudo-random samples from the standard half-normal distribution using a variation on the Box-Müller method proposed by Henze (1986) for the simulation of skew-normal variates. With these simulated samples we investigated the bias and mean squared error (MSE) of the MM and ML estimators of ξ and η identified in Section 2.3, and their bias-corrected counterparts given in Equations (2.4.2), (2.4.9) and (2.4.10). Throughout the study it was assumed that both parameters were unknown. For the estimators arising from the method of moments we also recorded the percentage of the estimates which were inadmissible. In addition to these summaries for the point estimates, we also quantified the actual coverages of the nominal 90%, 95% and 99%, MM based, confidence sets given by Equations (2.4.3)-(2.4.6) and their ML based analogues specified in Equations (2.4.11)-(2.4.13) and (2.4.15).

Tables 2.2 and 2.3 summarize the results obtained for the biases and MSEs of the various estimators of ξ and η, respectively. Also included in Table 2.2 are the observed percentages of inadmissible estimates of ξ for those estimators associated with the method of moments. From an inspection of Table 2.2 we conclude that

bias-correction leads to improved sampling properties of the point estimators of ξ, with both the bias and MSE being smaller than for the original estimators. Whilst $\tilde{\xi}_{BC}$ has the smallest bias of the four estimators considered, the bias-corrected ML based estimator, $\hat{\xi}_{BC}$, is to be preferred given its consistently smaller MSE. In addition, the fact that a high proportion of the values of $\tilde{\xi}_{BC}$ are inadmissible rules it out as a realistic competitor. We note that $\hat{\xi}_{BC}$ is marginally negatively biased.

Table 2.2 Empirical bias and MSE of MM and ML based estimators of ξ calculated from one million simulated samples of size n. The non-bias-corrected estimators, $\tilde{\xi}$ and $\hat{\xi}$, are denoted as MM(NBC) and ML(NBC), and their bias-corrected counterparts, $\tilde{\xi}_{BC}$ and $\hat{\xi}_{BC}$, as MM(BC) and ML(BC).

Sample size, n	Estimator	Bias	MSE	% inadmissible
20	MM(NBC)	0.03420	0.01754	41.68
	MM(BC)	-0.00005	0.01730	30.97
	ML(NBC)	0.05989	0.00688	
	ML(BC)	-0.00052	0.00355	
30	MM(NBC)	0.02280	0.01156	43.28
	MM(BC)	-0.00004	0.01146	34.62
	ML(NBC)	0.04042	0.00318	
	ML(BC)	-0.00035	0.00162	
40	MM(NBC)	0.01715	0.00861	44.15
	MM(BC)	0.00002	0.00855	36.65
	ML(NBC)	0.03064	0.00183	
	ML(BC)	-0.00012	0.00093	
50	MM(NBC)	0.01378	0.00690	44.87
	MM(BC)	0.00007	0.00686	38.17
	ML(NBC)	0.02459	0.00119	
	ML(BC)	-0.00011	0.00060	
100	MM(NBC)	0.00690	0.00342	46.37
	MM(BC)	0.00005	0.00341	41.67
	ML(NBC)	0.01242	0.00031	
	ML(BC)	-0.00002	0.00015	

As is evident from Table 2.3, bias-correction reduces both the bias and MSE of the ML based estimator of η. From the equivalent results for the MM based estimators we see that whilst the bias is reduced, the MSE is increased by bias-correction. Although the bias of the bias-corrected MM based estimator is the

smallest, its ML counterpart, $\hat{\eta}_{BC}$, wins out as its MSE is consistently the smallest. The latter estimator is marginally negatively biased.

Table 2.3 Empirical bias and MSE of MM and ML based estimators of η calculated from one million simulated samples of size n. The non-bias-corrected estimators, $\tilde{\eta}$ and $\hat{\eta}$, are denoted as MM(NBC) and ML(NBC), and their bias-corrected counterparts, $\tilde{\eta}_{BC}$ and $\hat{\eta}_{BC}$, as MM(BC) and ML(BC).

Sample size, n	Estimator	Bias	MSE
20	MM(NBC)	-0.04311	0.03574
	MM(BC)	-0.00018	0.03699
	ML(NBC)	-0.06091	0.02818
	ML(BC)	-0.03652	0.02709
30	MM(NBC)	-0.02867	0.02383
	MM(BC)	-0.00005	0.02438
	ML(NBC)	-0.04084	0.01813
	ML(BC)	-0.02444	0.01763
40	MM(NBC)	-0.02138	0.01787
	MM(BC)	0.00008	0.01819
	ML(NBC)	-0.03070	0.01334
	ML(BC)	-0.01835	0.01305
50	MM(NBC)	-0.01712	0.01433
	MM(BC)	0.00005	0.01453
	ML(NBC)	-0.02457	0.01053
	ML(BC)	-0.01467	0.01034
100	MM(NBC)	-0.00865	0.00717
	MM(BC)	-0.00006	0.00722
	ML(NBC)	-0.01245	0.00515
	ML(BC)	-0.00747	0.00510

Summarizing our findings from Tables 2.2 and 2.3, we identify the ML based bias-corrected pair $\hat{\xi}_{BC}$ and $\hat{\eta}_{BC}$ as having the best sampling properties of the various estimators considered.

In Table 2.4 we present the empirical coverages of the nominal 90%, 95% and 99% confidence intervals for ξ given by Equations (2.4.3), (2.4.4), (2.4.11) and (2.4.12). The standard error of any entry in this and the subsequent two tables is, at most, 0.03%. From these empirical coverage levels we see that, for a nominal level of 90%, the coverage of the MM based bias-corrected confidence interval specified by (2.4.4) is marginally closest to the nominal level. For the other two levels, the coverage of the ML based bias-corrected confidence interval given by (2.4.12) is closest. We note that whilst the coverages of all four intervals fall

marginally below the nominal level, they are nevertheless excellent. Indeed, the largest difference between the nominal and empirical levels is less than 1.9%, this for a sample size of only 20.

Table 2.4 Empirical coverage levels for nominal 90%, 95% and 99% MM and ML based, non-bias-corrected (NBC) and bias-corrected (BC), confidence intervals for ξ. The standard error of any entry is, at most, 0.03%.

Sample size, n	Method	Nominal level 90%	95%	99%
20	MM (NBC)	88.17	93.10	97.41
	MM(BC)	89.61	94.10	97.85
	ML (NBC)	88.90	94.20	98.66
	ML(BC)	89.19	94.41	98.74
30	MM (NBC)	88.81	93.76	97.97
	MM(BC)	89.76	94.42	98.24
	ML (NBC)	89.31	94.47	98.79
	ML(BC)	89.50	94.61	98.83
40	MM (NBC)	89.19	94.10	98.25
	MM(BC)	89.92	94.59	98.44
	ML (NBC)	89.48	94.65	98.86
	ML(BC)	89.62	94.74	98.90
50	MM (NBC)	89.26	94.24	98.38
	MM(BC)	89.84	94.63	98.53
	ML (NBC)	89.57	94.72	98.90
	ML(BC)	89.68	94.79	98.93
100	MM (NBC)	89.71	94.67	98.71
	MM(BC)	90.00	94.87	98.78
	ML (NBC)	89.82	94.86	98.96
	ML(BC)	89.87	94.89	98.97

Table 2.5 provides a similar summary for the empirical coverages of the confidence intervals for η given by Equations (2.4.5) and (2.4.13). From a consideration of its content we see that the coverage of the MM based interval consistently exceeds the nominal level, whilst that of its ML counterpart falls consistently short of it. For nominal levels of 90% and 95%, the ML based interval tends generally to hold the level best, whereas for one of 99% the two intervals hold the level equally well. Again, the empirical coverages of both intervals are excellent, even for samples of only 20 observations, with the largest disparity between the empirical and nominal levels being less than 1.2%.

Table 2.5 Empirical coverage levels for nominal 90%, 95% and 99% MM and ML based confidence intervals for η. The standard error of any entry is, at most, 0.03%.

Sample size, n	Method	Nominal level 90%	95%	99%
20	MM	91.19	95.77	99.10
	ML	89.53	94.67	98.88
30	MM	90.84	95.50	99.07
	ML	89.61	94.73	98.92
40	MM	90.64	95.43	99.08
	ML	89.73	94.78	98.93
50	MM	90.44	95.32	99.04
	ML	89.75	94.86	98.95
100	MM	90.22	95.12	99.02
	ML	89.78	94.88	98.96

Table 2.6 Empirical coverage levels for nominal 90%, 95% and 99% MM and ML based confidence regions for (ξ,η). The standard error of any entry is, at most, 0.03%.

Sample size, n	Method	Nominal level 90%	95%	99%
20	MM	90.87	95.32	98.86
	ML	90.48	95.32	99.09
30	MM	90.62	95.24	98.88
	ML	90.29	95.19	99.07
40	MM	90.48	95.20	98.92
	ML	90.22	95.17	99.05
50	MM	90.38	95.15	98.92
	ML	90.21	95.16	99.04
100	MM	90.17	95.07	98.96
	ML	90.05	95.04	99.02

The final table in this section, Table 2.6, provides a summary of the empirical coverage levels of the confidence regions for (ξ,η) associated with Equations (2.4.6) and (2.4.15). We note that, whilst the empirical coverages of both regions marginally exceed the nominal levels, the largest difference between the two, corresponding to a sample size of 20, is less than 0.9%. The ML based region

consistently outperforms its MM based analogue, although the differences between the observed coverages are very small indeed

2.6 Hypothesis Testing

Given the well-established equivalence between hypothesis testing and confidence set construction (Neyman, 1937; Lehmann, 1959, pp. 78-83; Aitchison, 1964, 1965), hypotheses concerning the parameters of the general half-normal distribution can be tested via the appropriate use of the large-sample confidence sets derived in Sections 2.4.1 and 2.4.2. The coverage results in Section 2.5 indicate that tests incorporating the likelihood theory based sets are generally to be preferred to their MM based counterparts. Those same empirical coverages also indicate that the use of the likelihood based confidence set associated with the pair (ξ,η) will result in a test that is marginally conservative. On the other hand, the analogous tests for ξ and η will have type I errors marginally in excess of their nominal levels.

2.7 An Illustrative Example: The Body Fat Data Revisited

In order to illustrate the methodology developed in the preceding sections, we present an analysis of the body fat measurements of the 102 male athletes considered previously in Section 1.4.4.2. The data values for this subsample are the ones underlined in Table 1.7 of Chapter 1. A histogram of the data is given in Figure 2.3.

As we saw in Section 1.4.4.2, if the general skew-normal distribution is fitted to these data then the resulting MM and ML solutions correspond to general half-normal distributions. Fitting the general half-normal distribution directly to these data, the bias-corrected ML based point estimates of the location and scale parameters are $\hat{\xi}_{BC} = 5.57$ and $\hat{\eta}_{BC} = 4.84$, respectively. Separate nominally 95% confidence intervals for ξ and η, calculated using (2.4.12) and (2.4.13) are (5.41, 5.63) and (4.25, 5.61), respectively. As we commented in Section 1.4.4.2, a percentage body fat measurement of 5% is generally considered in the sports physiology literature as being the lower threshold value for a healthy elite athlete. We note that the interval for the threshold parameter ξ does not include this value. Should we wish to force the threshold value of ξ to be some prespecified value then of course we could do so. We would then estimate η using $\hat{\eta}_{\xi}$, and obtain approximate confidence intervals for it using the distributional result (2.4.14).

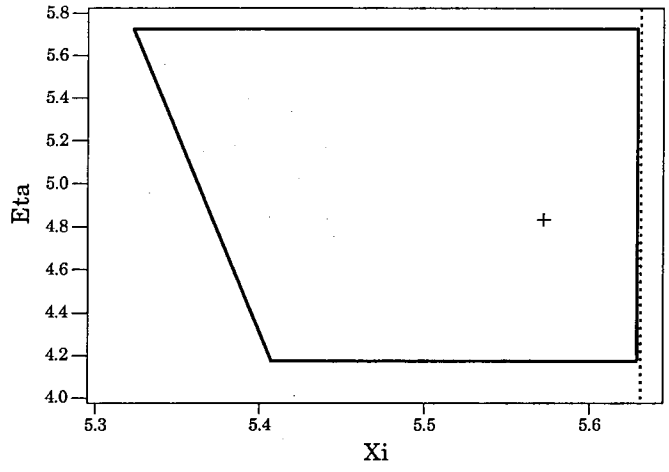

Figure 2.2 Approximate 95% confidence region for (ξ,η) for the percentage body fat measurements of the 102 elite male athletes. The cross represents $(\hat{\xi}_{BC},\hat{\eta}_{BC})$. The dashed vertical line has ξ co-ordinate $\hat{\xi} = y_{(1)} = 5.63$.

An approximate 95% confidence region for (ξ,η) obtained using (2.4.15) is given in Figure 2.2. Also represented in this figure is the position of the bias-corrected ML solution $(\hat{\xi}_{BC},\hat{\eta}_{BC})$. The (ξ,η) co-ordinates of the extreme points defining the displayed confidence region are: (5.63, 5.73), (5.63, 4.18), (5.32, 5.73) and (5.41, 4.18). Given these values, any hypothesis concerning the two parameters ξ and η which specified a value as low as 5 for the threshold parameter ξ would be rejected by a two-sided test with a nominal significance level of 5%. For illustrative purposes, in Figure 2.3 the densities corresponding to the two extreme points (5.63, 4.18) and (5.32, 5.73) have been superimposed upon the histogram of the data together with the density of the bias-corrected ML solution (5.57, 4.84). The densities corresponding to the other two extreme points, (5.63, 5.73) and (5.41, 4.18), are very similar to those for (5.32, 5.73) and (5.63, 4.18), respectively, and have been omitted so as not to obscure the content of the plot. Visually the density of the bias-corrected ML solution provides a reasonable fit, an impression confirmed by the chi-squared goodness-of-fit test. The value of the chi-squared statistic, based on eight class intervals with expected frequencies in excess of 5 and formed from those used in the histogram of Figure 2.3, corresponded to a p-value of around 0.15. Nevertheless, the disparities between the fitted densities and the observed frequencies evident in Figure 2.3 suggest that perhaps a heavier-tailed

distribution, such as the location and scale extension of a half-t distribution, might provide an even better fit to the data.

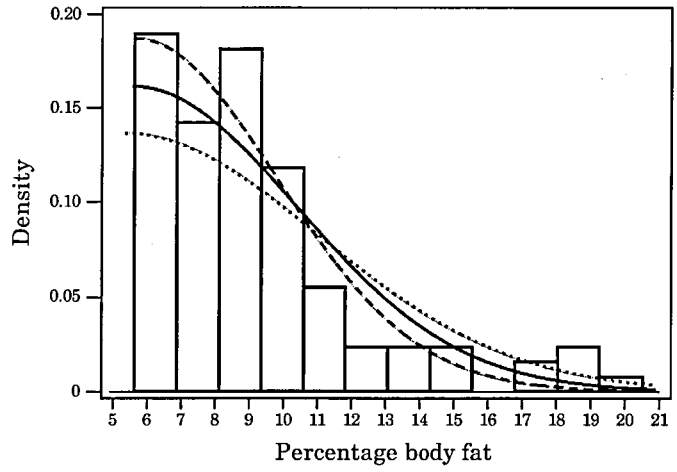

Figure 2.3 Histogram of the percentage body fat measurements of the 102 elite male athletes with superimposed half-normal densities: ——, bias-corrected maximum likelihood solution (5.57, 4.84); – – –, (5.63, 4.18); · · · · ·, (5.32, 5.73).

2.8 Summary and Directions for Future Research

In this, the last, section of the present chapter, we summarize the content and conclusions of the previous sections, and identify related issues for potential future research.

2.8.1 Summary

In Section 2.2 we extended the standard half-normal distribution via the inclusion of location and scale parameters, by so doing producing the general half-normal distribution. In passing, we noted the connections between the derived distribution and other well known distributions. Point estimates of the distribution's parameters arising from the method of moments and the method of maximum likelihood were identified in Section 2.3. Asymptotic distributional results for both types of estimator were derived in Section 2.4. Using these results, we established alternative bias-corrected estimators and derived constructions for large-sample confidence sets.

In Section 2.5 we presented the results of a Monte Carlo simulation study conducted in order to explore the sampling properties of the estimators and the coverages of the confidence sets derived in Sections 2.3 and 2.4. From this study we

concluded that the bias-corrected ML based point estimators had superior sampling properties, their MSEs being consistently the smallest. Regarding the coverages of the various confidence sets, bias-correction was found to improve the empirical coverage levels. The bias-corrected ML based interval for ξ given by (2.4.12) was found to have superior coverage for the nominal levels of 95% and 99%, whilst its MM based counterpart given by (2.4.4) held the 90% nominal level best. For all three nominal levels considered, the empirical coverage of the interval with best coverage fell marginally short of the nominal level.

Regarding the confidence intervals for η, and confidence regions for (ξ,η), the ML based confidence sets given by (2.4.13) and (2.4.15) were identified as having superior coverages to their MM based analogues specified in (2.4.5) and (2.4.6). The empirical coverage of the interval for η specified in (2.4.13) was found to fall marginally below the nominal level, whilst that for the confidence region for (ξ,η) given by (2.4.13) fell marginally above the nominal level. For the confidence sets for ξ, η and (ξ,η) with best coverages, the single largest disparity between the empirical and nominal coverage levels was found to be just 0.48%, this difference corresponding to the confidence region for (ξ,η) and a sample size of just 20. Overall, the coverages of the confidence sets with best coverages were classified as being excellent.

In Section 2.6 we briefly discussed the use of the derived large-sample confidence sets as a basis upon which to carry out hypothesis testing. In Section 2.7 we illustrated the methodology developed in the preceeding sections, obtaining point estimates and confidence sets for the parameters of an assumed half-normal parent population fitted to the body fat measurements of elite male athletes.

2.8.2 Directions for Future Research

As we explained in the introductory section of this chapter, our primary motivation for developing methods of inference for the general half-normal distribution was the distribution's importance as a limiting distribution of the skew-normal class. Another, perhaps more important, reason for considering such methods is the potential relevance of the distribution in the modelling of real skew data. However, the general half-normal distribution is not the only one obtained by folding a symmetric parent population about its centre which might prove useful to modellers of skew data. As we commented at the end of our illustrative analysis of

Section 2.7, for highly skew data exhibiting a sharp cut-off to one side of the mode, and a heavy tail to the other, the generalization of the half-t distribution obtained on introducing location and scale parameters to its definition, could well provide a better model than the general half-normal distribution considered here. To our knowledge, this extension of the half-t distribution has not previously been proposed in the literature. We note that the basic distributional properties of the half-t distribution were studied by Psarakis & Panaretos (1990), who incorrectly refer to it as the "folded" t distribution.

Our treatment regarding inference for the parameters of the general half-normal distribution was founded upon two of the most commonly utilized methods of estimation; namely, the method of moments and the method of maximum likelihood. It would be of interest to compare the results we have obtained here with those arising from other methods of estimation. As the results of our simulation experiment showed, the coverages of the (bias-corrected) ML based confidence sets arising from asymptotic theory are excellent, even for sample sizes as small as 20. It would be interesting to compare their coverages for even smaller samples with those of confidence sets obtained, say, on applying the non-parametric and parametric versions of the bootstrap. General bootstrap approaches to confidence region construction are discussed in Davison & Hinkley (1997, Chapters 3 & 5). An alternative approach to constructing bootstrap confidence regions, based on Tukey's depth criterion, is considered by Yeh & Singh (1997). A related issue of interest is that of the minimization of the area of any given confidence region for a specified confidence level. The paper by Arnold & Shavelle (1998) provides a useful introductory reference on this point.

In this chapter we have considered inference for the most basic of set-ups, that arising from random sampling. Amongst other scenarios which could be addressed, we mention those of regression and time series modelling. The regression problem involving half-normal errors with $\xi = 0$ is one which has been considered in the stochastic frontier modelling literature by Aigner & Chu (1968), Schmidt (1976) and Aigner *et al.* (1977).

Clearly, multivariate extensions of the general half-normal distribution are possible. For any such model, the threshold parameters included in its definition will play a crucial role in terms of the distribution's potential relevance to the modelling of skew multivariate data.

PART II

CONTRIBUTIONS TO THE ANALYSIS OF SKEW DATA ON THE CIRCLE

Chapter 3 The Large-sample Distribution of Key Circular Statistics

3.1 Introduction

In this chapter we consider the joint distribution of four fundamental statistics used in the analysis of circular data: the mean direction, the mean resultant length, the second central sine moment and the second central cosine moment. Between them, these statistics provide summaries of location, concentration, skewness and kurtosis of a sample of circular data.

The definitions of the four statistics, and their analogous population measures, are reviewed in Section 3.2. In Section 3.3, we use Taylor expansion to derive the asymptotic joint distribution of the statistics, in so doing simplifying and extending the results of Mardia (1972, Section 4.9.2). Our motivation for presenting these results is their importance in the modelling of skew distributed circular data, an area largely ignored in a literature dominated by a tacit, although often unrealistic, assumption of underlying symmetry. On the basis of the results in Section 3.3, in Section 3.4 we define bias-corrected estimators for the population measures.

Large-sample inference for the population measures is addressed in Section 3.5. We give new constructions for confidence intervals for the individual measures as well as those for confidence sets for pairs, etc. of the measures. The various constructions proposed allow for bias-correction as well as for an assumption of underlying symmetry.

In Section 3.6 we discuss the problems which can arise when applying the various inferential tools. In order to illustrate the use of the developed methodology and some of the problems associated with its application, in Section 3.7 we present analyses of three data sets recorded during animal orientation experiments.

The chapter closes with Section 3.8, in which we provide a summary of the Chapter's content and an indication of related lines of potential future research.

3.2 Fundamental Statistics for Circular Data

Let θ denote a random angle and $\theta_1,...,\theta_n$ a random sample of n observations from a circular distribution with characteristic function $\{\psi_p : p = 0,\pm1,...\}$. Then $\psi_p = \alpha_p + i\beta_p$ where $\alpha_p = E(\cos p\theta)$ and $\beta_p = E(\sin p\theta)$ denote the trigonometric moments of θ about the zero direction. For $p = 1$, $\psi_1 = \alpha_1 + i\beta_1 = \rho e^{i\mu}$, where μ is the mean direction and ρ the mean resultant length. The trigonometric moments about the mean direction are given by

$$\bar{\alpha}_p = E\{\cos p(\theta - \mu)\}, \quad \bar{\beta}_p = E\{\sin p(\theta - \mu)\}.$$

The method of moments estimator of ρ is $\bar{R} = \left(a_1^2 + b_1^2\right)^{1/2}$, where $a_p = \frac{1}{n}\sum_{i=1}^{n}\cos p\,\theta_i$ and $b_p = \frac{1}{n}\sum_{i=1}^{n}\sin p\,\theta_i$ are the sample trigonometric moments about the zero direction. If $\bar{R} = 0$, the method of moments estimator of μ, $\bar{\theta}$, is undefined. If $\bar{R} > 0$,

$$\bar{\theta} = \begin{cases} \tan^{-1}\left(b_1/a_1\right) & \text{if } a_1 \geq 0 \\ \pi + \tan^{-1}\left(b_1/a_1\right) & \text{if } a_1 < 0 \end{cases}$$

where $\tan^{-1}(x) \in [-\pi/2, \pi/2]$, for x a real.

Mardia (1972, Section 3.7) introduced standardized versions of $\bar{\beta}_2$ and $\bar{\alpha}_2$ as circular measures of population skewness and kurtosis, respectively. Denoting these measures by s and k,

$$s = \frac{\bar{\beta}_2}{v^{3/2}} = \frac{\bar{\beta}_2}{(1-\rho)^{3/2}} \quad \text{and} \quad k = \frac{\bar{\alpha}_2 - (1-v)^4}{v^2} = \frac{\bar{\alpha}_2 - \rho^4}{(1-\rho)^2},$$

$v = 1 - \rho$ denoting the circular variance of θ. The standardizations used in the definitions of s and k were motivated by considerations of the properties of the two measures for concentrated distributions on the circle. For a symmetric distribution, both $\bar{\beta}_2$ and s equal zero. The standardization used in the definition of k ensures that $k = 0$ for the wrapped normal distribution. See Mardia (1972, pp. 74-76) and Mardia & Jupp (1999, pp. 21-22). However, s and k are not in general scale invariant, a consequence of there existing no standardization transform for circular data equivalent to that routinely used for data on the line. For the same reason, $\bar{\beta}_2$ and $\bar{\alpha}_2$ are not generally scale invariant either, but general expressions for the

large-sample properties of their sample analogues, $\bar{b}_2 = \frac{1}{n}\sum_{i=1}^{n} \sin 2(\theta_i - \bar{\theta})$ and $\bar{a}_2 = \frac{1}{n}\sum_{i=1}^{n} \cos 2(\theta_i - \bar{\theta})$, are relatively easy to derive. We note that $\bar{\beta}_2$ was first proposed as a measure of circular skewness by Batschelet (1965), and the option of using the fundamental measures $\bar{\beta}_2$ and $\bar{\alpha}_2$, rather than s and k, was referred to by Mardia (1972, p. 76).

3.3 The Asymptotic Distribution of $(\bar{\theta}, \bar{R}, \bar{b}_2, \bar{a}_2)$

Mardia (1972, Section 4.9) derived asymptotic results for the general form of the joint distribution of $(\bar{\theta}, \bar{R})$. He also gave the joint distribution of the sample analogues of s and k, $\tilde{s}$ and $\tilde{k}$, for a parent population which is symmetric about the zero direction. In this section, we simplify and extend Mardia's results, providing concise general results for the full asymptotic joint distribution of $(\bar{\theta}, \bar{R}, \bar{b}_2, \bar{a}_2)$. The main result of the chapter is summarized in the following theorem.

Theorem 3.1 *Let $\theta_1, ..., \theta_n$ be a random sample of n observations from a circular distribution with $\rho \in (0, 1)$. Then the asymptotic joint distribution of $(\bar{\theta}, \bar{R}, \bar{b}_2, \bar{a}_2)$ is multivariate normal with*

$$E(\bar{\theta}) = \mu - \frac{\bar{\beta}_2}{2n\rho^2} + O(n^{-3/2}), \quad E(\bar{R}) = \rho + \frac{(1-\bar{\alpha}_2)}{4n\rho} + O(n^{-3/2}),$$

$$E(\bar{b}_2) = \bar{\beta}_2 + \frac{1}{n}\left(-\frac{\bar{\beta}_3}{\rho} - \frac{\bar{\beta}_2}{\rho^2} + \frac{2\bar{\alpha}_2\bar{\beta}_2}{\rho^4}\right) + O(n^{-3/2}), \tag{3.3.1}$$

$$E(\bar{a}_2) = \bar{\alpha}_2 + \frac{1}{n}\left\{1 - \frac{\bar{\alpha}_3}{\rho} - \frac{\bar{\alpha}_2(1-\bar{\alpha}_2) + \bar{\beta}_2^2}{\rho^2}\right\} + O(n^{-3/2}).$$

$$\mathrm{var}(\bar{\theta}) = \frac{(1-\bar{\alpha}_2)}{2n\rho^2} + O(n^{-3/2}), \quad \mathrm{var}(\bar{R}) = \frac{(1-2\rho^2+\bar{\alpha}_2)}{2n} + O(n^{-3/2}),$$

$$\mathrm{var}(\bar{b}_2) = \frac{1}{n}\left[\frac{1-\bar{\alpha}_4}{2} - 2\bar{\alpha}_2 - \bar{\beta}_2^2 + \frac{2\bar{\alpha}_2}{\rho}\left\{\bar{\alpha}_3 + \frac{\bar{\alpha}_2(1-\bar{\alpha}_2)}{\rho}\right\}\right] + O(n^{-3/2}), \tag{3.3.2}$$

$$\mathrm{var}(\bar{a}_2) = \frac{1}{n}\left[\frac{1+\bar{\alpha}_4}{2} - \bar{\alpha}_2^2 + \frac{2\bar{\beta}_2}{\rho}\left\{\bar{\beta}_3 + \frac{\bar{\beta}_2(1-\bar{\alpha}_2)}{\rho}\right\}\right] + O(n^{-3/2}).$$

$$
\begin{aligned}
\operatorname{cov}(\bar{\theta},\bar{R}) &= \frac{\bar{\beta}_2}{2n\rho} + O(n^{-3/2}),\\
\operatorname{cov}(\bar{\theta},\bar{b}_2) &= \frac{1}{n}\left\{\frac{1}{2} - \frac{\bar{\alpha}_3}{2\rho} - \frac{\bar{\alpha}_2(1-\bar{\alpha}_2)}{\rho^2}\right\} + O(n^{-3/2}),\\
\operatorname{cov}(\bar{\theta},\bar{a}_2) &= \frac{1}{n}\left\{\frac{\bar{\beta}_3}{2\rho} + \frac{\bar{\beta}_2(1-\bar{\alpha}_2)}{\rho^2}\right\} + O(n^{-3/2}),\\
\operatorname{cov}(\bar{R},\bar{b}_2) &= \frac{1}{n}\left(-\rho\bar{\beta}_2 + \frac{\bar{\beta}_3}{2} - \frac{\bar{\alpha}_2\bar{\beta}_2}{\rho}\right) + O(n^{-3/2}),\\
\operatorname{cov}(\bar{R},\bar{a}_2) &= \frac{1}{n}\left\{\frac{\rho(1-2\bar{\alpha}_2)}{2} + \frac{\bar{\alpha}_3}{2} + \frac{\bar{\beta}_2^2}{\rho}\right\} + O(n^{-3/2}),\\
\operatorname{cov}(\bar{b}_2,\bar{a}_2) &= \frac{1}{n}\left\{(1-\bar{\alpha}_2)\bar{\beta}_2 + \frac{\bar{\beta}_4}{2} - \frac{\bar{\alpha}_2\bar{\beta}_3 + \bar{\beta}_2\bar{\alpha}_3}{\rho} - \frac{2\bar{\alpha}_2\bar{\beta}_2(1-\bar{\alpha}_2)}{\rho^2}\right\}\\
&\qquad + O(n^{-3/2}),
\end{aligned}
\tag{3.3.3}
$$

if the terms of $O(n)$ for the variances of $\bar{b}_2$ and $\bar{a}_2$ in (3.3.2) are positive.

Proof Proceeding as in Mardia (1972, Section 4.9.2), since $\rho \neq 0$ we can assume that after a suitable rotation $\alpha_1 > 0$, so that, with probability 1, $\bar{\theta} = \tan^{-1}(b_1/a_1)$. In addition, the other three statistics, $\bar{R} = (a_1^2 + b_1^2)^{1/2}$, $\bar{b}_2 = \dfrac{(a_1^2 - b_1^2)b_2 - 2a_1 b_1 a_2}{a_1^2 + b_1^2}$ and $\bar{a}_2 = \dfrac{(a_1^2 - b_1^2)a_2 + 2a_1 b_1 b_2}{a_1^2 + b_1^2}$, are also differentiable functions of a_1, b_1, a_2, b_2.

Appealing to the central limit theorem, the asymptotic distribution of $(a_1, b_1, a_2, b_2)^T$ is $N(\underline{\xi}, \underline{\Sigma})$ where $\underline{\xi} = (\xi_1, \xi_2, \xi_3, \xi_4)^T = (\alpha_1, \beta_1, \alpha_2, \beta_2)^T$ and $\underline{\Sigma} = (\sigma_{ij})$, with

$$
\begin{aligned}
\sigma_{11} &= \operatorname{var}(a_1) = (1+\alpha_2 - 2\alpha_1^2)/2n,\ \sigma_{22} = \operatorname{var}(b_1) = (1-\alpha_2 - 2\beta_1^2)/2n,\\
\sigma_{33} &= \operatorname{var}(a_2) = (1+\alpha_4 - 2\alpha_2^2)/2n,\ \sigma_{44} = \operatorname{var}(b_2) = (1-\alpha_4 - 2\beta_2^2)/2n,\\
\sigma_{12} &= \sigma_{21} = \operatorname{cov}(a_1, b_1) = (\beta_2 - 2\alpha_1\beta_1)/2n,\\
\sigma_{13} &= \sigma_{31} = \operatorname{cov}(a_1, a_2) = (\alpha_1 + \alpha_3 - 2\alpha_1\alpha_2)/2n,\\
\sigma_{14} &= \sigma_{41} = \operatorname{cov}(a_1, b_2) = (\beta_1 + \beta_3 - 2\alpha_1\beta_2)/2n,\\
\sigma_{23} &= \sigma_{32} = \operatorname{cov}(b_1, a_2) = (\beta_3 - \beta_1 - 2\beta_1\alpha_2)/2n,\\
\sigma_{24} &= \sigma_{42} = \operatorname{cov}(b_1, b_2) = (\alpha_1 - \alpha_3 - 2\beta_1\beta_2)/2n,\\
\sigma_{34} &= \sigma_{43} = \operatorname{cov}(a_2, b_2) = (\beta_4 - 2\alpha_2\beta_2)/2n.
\end{aligned}
\tag{3.3.4}
$$

Now apply Lemma 1.2 of Chapter 1. For the functions $h_1 = \tan^{-1}(\beta_1/\alpha_1)$, $h_2 = (\alpha_1^2+\beta_1^2)^{1/2}$, $h_3 = \dfrac{(\alpha_1^2-\beta_1^2)\beta_2 - 2\alpha_1\beta_1\alpha_2}{\alpha_1^2+\beta_1^2}$ and $h_4 = \dfrac{(\alpha_1^2-\beta_1^2)\alpha_2 + 2\alpha_1\beta_1\beta_2}{\alpha_1^2+\beta_1^2}$ we obtain the following non-zero partial derivatives, those for h_3 and h_4 having been simplified using the identities in (3.3.5):

$$h_1^{(1)} = -\beta_1/\rho^2,\ h_1^{(2)} = \alpha_1/\rho^2,$$
$$h_1^{(11)} = 2\alpha_1\beta_1/\rho^4,\ h_1^{(22)} = -2\alpha_1\beta_1/\rho^4,\ h_1^{(12)} = -(\alpha_1^2-\beta_1^2)/\rho^4.$$

$$h_2^{(1)} = \alpha_1/\rho,\ h_2^{(2)} = \beta_1/\rho,$$
$$h_2^{(11)} = \beta_1^2/\rho^3,\ h_2^{(22)} = \alpha_1^2/\rho^3,\ h_2^{(12)} = -\alpha_1\beta_1/\rho^3.$$

$$h_3^{(1)} = 2\beta_1\bar{\alpha}_2/\rho^2,\ h_3^{(2)} = -2\alpha_1\bar{\alpha}_2/\rho^2,$$
$$h_3^{(3)} = -2\alpha_1\beta_1/\rho^2,\ h_3^{(4)} = (\alpha_1^2-\beta_1^2)/\rho^2,$$
$$h_3^{(11)} = \{4\beta_1(\beta_1\beta_2+\alpha_1\alpha_2) - 8\alpha_1\beta_1\bar{\alpha}_2\}/\rho^4,$$
$$h_3^{(12)} = \{8\alpha_1\beta_1\beta_2 + 2\alpha_2(\alpha_1^2-3\beta_1^2) - 8\beta_1^2\bar{\alpha}_2\}/\rho^4,$$
$$h_3^{(13)} = 2\beta_1(\alpha_1^2-\beta_1^2)/\rho^4,\ h_3^{(14)} = 4\alpha_1\beta_1^2/\rho^4,$$
$$h_3^{(22)} = \{4\alpha_1(\beta_1\alpha_2-\alpha_1\beta_2) + 8\alpha_1\beta_1\bar{\alpha}_2\}/\rho^4,$$
$$h_3^{(23)} = -2\alpha_1(\alpha_1^2-\beta_1^2)/\rho^4,\ h_3^{(24)} = -4\alpha_1^2\beta_1/\rho^4.$$

$$h_4^{(1)} = -2\beta_1\bar{\beta}_2/\rho^2,\ h_4^{(2)} = 2\alpha_1\bar{\beta}_2/\rho^2,$$
$$h_4^{(3)} = (\alpha_1^2-\beta_1^2)/\rho^2,\ h_4^{(4)} = 2\alpha_1\beta_1/\rho^2,$$
$$h_4^{(11)} = 4\beta_1(\alpha_1\bar{\beta}_2 - \beta_1\bar{\alpha}_2)/\rho^4,\ h_4^{(12)} = \{4\alpha_1\beta_1\bar{\alpha}_2 - 2(\alpha_1^2-\beta_1^2)\bar{\beta}_2\}/\rho^4,$$
$$h_4^{(13)} = 4\alpha_1\beta_1^2/\rho^4,\ h_4^{(14)} = -2\beta_1(\alpha_1^2-\beta_1^2)/\rho^4,$$
$$h_4^{(22)} = -4\alpha_1(\alpha_1\bar{\alpha}_2 + \beta_1\bar{\beta}_2)/\rho^4,\ h_4^{(23)} = -4\alpha_1^2\beta_1/\rho^4,\ h_4^{(24)} = 2\alpha_1(\alpha_1^2-\beta_1^2)/\rho^4.$$

In these expressions, $h_k^{(i)} = \partial h_k/\partial\xi_i$ and $h_k^{(ij)} = \partial^2 h_k/\partial\xi_i\,\partial\xi_j$. Using these partial derivatives together with the variances and covariances given in (3.3.4), and the identities

$$\begin{aligned}
\bar{\alpha}_2\rho^2 &= (\alpha_1^2-\beta_1^2)\alpha_2 + 2\alpha_1\beta_1\beta_2,\\
\bar{\beta}_2\rho^2 &= (\alpha_1^2-\beta_1^2)\beta_2 - 2\alpha_1\beta_1\alpha_2,\\
\bar{\alpha}_3\rho^3 &= \alpha_1(\alpha_1^2-3\beta_1^2)\alpha_3 + \beta_1(3\alpha_1^2-\beta_1^2)\beta_3,\\
\bar{\beta}_3\rho^3 &= \alpha_1(\alpha_1^2-3\beta_1^2)\beta_3 - \beta_1(3\alpha_1^2-\beta_1^2)\alpha_3,\\
\bar{\alpha}_4\rho^4 &= (\alpha_1^4-6\alpha_1^2\beta_1^2+\beta_1^4)\alpha_4 + 4\alpha_1\beta_1(\alpha_1^2-\beta_1^2)\beta_4,\\
\bar{\beta}_4\rho^4 &= (\alpha_1^4-6\alpha_1^2\beta_1^2+\beta_1^4)\beta_4 - 4\alpha_1\beta_1(\alpha_1^2-\beta_1^2)\alpha_4,
\end{aligned} \qquad (3.3.5)$$

we obtain, after lengthy but basic algebraic simplification, the expectations, variances and covariances given in (3.3.1)-(3.3.3). Asymptotic normality follows from Lemma 1.2 due to the asymptotic normality of $\left(a_1, b_1, a_2, b_2\right)^T$, the fact that the leading terms of the expectations in (3.3.1) are finite (as $\rho \neq 0$) and the assumption that the terms of order $O(n)$ for the variances of $\bar{b}_2$ and $\bar{a}_2$ in (3.3.2) are positive.

The bias terms in the expectations of $\bar{\theta}$ and $\bar{R}$ given in (3.3.1) are new to the literature. The formulae for the variances of $\bar{\theta}$ and $\bar{R}$, and their covariance, are equivalent but more concise than the expressions of Mardia (1972, p.110) given in terms of trigonometric moments about the zero direction. They also highlight the importance of the assumption that $\rho \neq 1$. The other expressions in (3.3.1)-(3.3.3) are also new to the literature.

Theorem 3.1 specifically excludes two important cases. The case for which $\rho = 0$ includes the uniform distribution as well as distributions which are multimodal and reflectively symmetric, or cyclically symmetric, or both. The analysis of data from multimodal distributions of these types is discussed in Section 3.6 and illustrated in Section 3.7.3. Also relevant in this context is the work of Mardia (1972, Section 4.9.2) in which the joint distribution of $\left(\bar{\theta}, \bar{R}\right)$ for a parent population with $\rho = 0$ is considered. The case for which $\rho = 1$ is pathological, corresponding to a point distribution.

3.4 Bias-corrected Estimators

Given the expectations in (3.3.1), we can define estimators of μ, ρ, $\bar{\beta}_2$, and $\bar{\alpha}_2$ corrected for bias to $O\left(n^{-1}\right)$. Replacing the population measures in the bias terms of (3.3.1) by their sample analogues, we obtain:

$$\begin{aligned}
\hat{\mu}_{BC} &= \bar{\theta} + \frac{\bar{b}_2}{2n\bar{R}^2}, \\
\hat{\rho}_{BC} &= \bar{R} - \frac{(1-\bar{a}_2)}{4n\bar{R}}, \\
\hat{\bar{\beta}}_{2,BC} &= \bar{b}_2 - \frac{1}{n}\left(-\frac{\bar{b}_3}{\bar{R}} - \frac{\bar{b}_2}{\bar{R}^2} + \frac{2\bar{a}_2\bar{b}_2}{\bar{R}^4}\right), \\
\hat{\bar{\alpha}}_{2,BC} &= \bar{a}_2 - \frac{1}{n}\left\{1 - \frac{\bar{a}_3}{\bar{R}} - \frac{\bar{a}_2(1-\bar{a}_2)+\bar{b}_2^2}{\bar{R}^2}\right\},
\end{aligned} \tag{3.4.1}$$

where $\bar{b}_p = \frac{1}{n}\sum_{i=1}^{n} \sin p(\theta_i - \bar{\theta})$ and $\bar{a}_p = \frac{1}{n}\sum_{i=1}^{n} \cos p(\theta_i - \bar{\theta})$.

If the parent population is assumed to be symmetric then, as the central sine moments are all zero, we would use the original estimator $\bar{\theta}$ in place of $\hat{\mu}_{BC}$. Clearly, if we assume the underlying distribution to be symmetric then there is no need to estimate $\bar{\beta}_2$. Also, under the assumption of symmetry, we would set $\bar{b}_2$ equal to zero in the expression for $\hat{\bar{\alpha}}_{2,BC}$.

3.5 Large-sample Inference

In this section, we use the asymptotic distributional results presented in Section 3.3 to derive large-sample confidence sets for the population measures μ, ρ, $\bar{\beta}_2$, and $\bar{\alpha}_2$. Hypothesis tests involving these measures can be based on the derived confidence sets as discussed in Section 2.6.

3.5.1 Confidence Intervals for μ

Using the results for $E(\bar{\theta})$ and $\text{var}(\bar{\theta})$ given in Theorem 3.1, and replacing ρ, $\bar{\alpha}_2$ and $\bar{\beta}_2$ by their sample analogues, $\bar{R}$, $\bar{a}_2$ and $\bar{b}_2$, as appropriate, then, including only those variance terms of $O(n^{-1})$, an approximate $100(1-\alpha)\%$, bias-corrected confidence interval for μ is given by

$$\bar{\theta} + \frac{\bar{b}_2}{2n\bar{R}^2} \pm z_{\alpha/2}\left(\frac{1-\bar{a}_2}{2n\bar{R}^2}\right)^{1/2}. \tag{3.5.1}$$

If the underlying distribution is assumed to be symmetric or, equivalently, we drop the bias term in (3.5.1), then the interval reduces to

$$\bar{\theta} \pm z_{\alpha/2}\left(\frac{1-\bar{a}_2}{2n\bar{R}^2}\right)^{1/2}. \tag{3.5.2}$$

This compares with the corresponding interval from Fisher & Lewis (1983) of

$$\bar{\theta} \pm \sin^{-1}\left\{z_{\alpha/2}\left(\frac{1-\bar{a}_2}{2n\bar{R}^2}\right)^{1/2}\right\}. \tag{3.5.3}$$

As $\sin^{-1}(x) \approx x$ for small x, then, for moderate sized samples from all but the most disperse of unimodal populations, we would expect to find little difference between the results obtained using (3.5.2) and (3.5.3).

3.5.2 Confidence Intervals for ρ

Similarly, an approximate $100(1-\alpha)\%$, bias-corrected, confidence interval for ρ is given by

$$\bar{R} - \frac{\left(1-\bar{a}_2\right)}{4n\bar{R}} \pm z_{\alpha/2}\left(\frac{1-2\bar{R}^2+\bar{a}_2}{2n}\right)^{1/2}. \tag{3.5.4}$$

Omitting the bias term we obtain the interval

$$\bar{R} \pm z_{\alpha/2}\left(\frac{1-2\bar{R}^2+\bar{a}_2}{2n}\right)^{1/2}, \tag{3.5.5}$$

which is effectively that which follows from the results of Mardia (1972, Section 4.9).

3.5.3 Confidence Regions for (μ, ρ)

An approximate $100(1-\alpha)\%$, bias-corrected, confidence region for $\underline{\xi}_{12} = (\mu, \rho)^T$ is given by those $\underline{\xi}_{12}$ for which

$$\left(\tilde{\underline{\xi}}_{12} - \underline{\xi}_{12}\right)^T \tilde{\underline{\Sigma}}_{12}^{-1}\left(\tilde{\underline{\xi}}_{12} - \underline{\xi}_{12}\right) \leq \chi^2_{2,1-\alpha}, \tag{3.5.6}$$

where $\chi^2_{2,1-\alpha}$ is the $(1-\alpha)$ quantile of the chi-squared distribution with two degrees of freedom,

$$\tilde{\underline{\xi}}_{12} = \left(\bar{\theta} + \frac{\bar{b}_2}{2n\bar{R}^2}, \bar{R} - \frac{\left(1-\bar{a}_2\right)}{4n\bar{R}}\right)^T \tag{3.5.7}$$

and $\tilde{\Sigma}_{12}=(\tilde{\sigma}_{ij})$ with $\tilde{\sigma}_{11}=\tilde{\text{var}}(\bar{\theta})=\left(\frac{1-\bar{a}_2}{2n\bar{R}^2}\right)$, $\tilde{\sigma}_{12}=\tilde{\text{cov}}(\bar{\theta},\bar{R})=\left(\frac{\bar{b}_2}{2n\bar{R}}\right)$ and $\tilde{\sigma}_{22}=\tilde{\text{var}}(\bar{R})=\left(\frac{1-2\bar{R}^2+\bar{a}_2}{2n}\right)$. An analogous, non-bias-corrected, confidence region follows similarly on omitting the two bias terms from (3.5.7). For a parent population which is assumed to be symmetric, the bias associated with $\bar{\theta}$ should be dropped from (3.5.7) and $\tilde{\sigma}_{12}$ set equal to zero.

3.5.4 Confidence Sets for $\bar{\beta}_2$ and $\bar{\alpha}_2$

Proceeding as in Sections 3.5.1 and 3.5.2, an approximate $100(1-\alpha)\%$, bias-corrected confidence interval for $\bar{\beta}_2$ is given by

$$\bar{b}_2-\tilde{\text{bias}}(\bar{b}_2)\pm z_{\alpha/2}\{\tilde{\text{var}}(\bar{b}_2)\}^{1/2}, \tag{3.5.8}$$

where $\tilde{\text{bias}}(\bar{b}_2)=\frac{1}{n}\left(-\frac{\bar{b}_3}{\bar{R}}-\frac{\bar{b}_2}{\bar{R}^2}+\frac{2\bar{a}_2\bar{b}_2}{\bar{R}^4}\right)$ and

$$\tilde{\text{var}}(\bar{b}_2)=\frac{1}{n}\left[\frac{1-\bar{a}_4}{2}-2\bar{a}_2-\bar{b}_2^2+\frac{2\bar{a}_2}{\bar{R}}\left\{\bar{a}_3+\frac{\bar{a}_2(1-\bar{a}_2)}{\bar{R}}\right\}\right].$$

The non-bias-corrected counterpart of this confidence interval follows on omitting the estimate of the bias from (3.5.8). Of course, for an underlying distribution which is assumed to be symmetric, $\bar{\beta}_2$ is identically zero.

Similarly, an approximate $100(1-\alpha)\%$, bias-corrected confidence interval for $\bar{\alpha}_2$ is given by

$$\bar{a}_2-\tilde{\text{bias}}(\bar{a}_2)\pm z_{\alpha/2}\{\tilde{\text{var}}(\bar{a}_2)\}^{1/2}, \tag{3.5.9}$$

where $\tilde{\text{bias}}(\bar{a}_2)=\frac{1}{n}\left\{1-\frac{\bar{a}_3}{\bar{R}}-\frac{\bar{a}_2(1-\bar{a}_2)+\bar{b}_2^2}{\bar{R}^2}\right\}$ and

$$\tilde{\text{var}}(\bar{a}_2)=\frac{1}{n}\left[\frac{1+\bar{a}_4}{2}-\bar{a}_2^2+\frac{2\bar{b}_2}{\bar{R}}\left\{\bar{b}_3+\frac{\bar{b}_2(1-\bar{a}_2)}{\bar{R}}\right\}\right].$$

The non-bias-corrected version of this confidence interval follows on omitting the estimate of the bias from (3.5.9). If the parent population is assumed to be

symmetric then the terms involving $\bar{b}_2$ should be set equal to zero in the expressions for $\widetilde{\text{bias}}(\bar{a}_2)$ and $\widetilde{\text{var}}(\bar{a}_2)$.

By analogy with the content of Section 3.5.3, an approximate $100(1-\alpha)\%$, bias-corrected, confidence region for $\underline{\xi}_{34} = (\bar{\beta}_2, \bar{\alpha}_2)^T$ is given by those $\underline{\xi}_{34}$ for which

$$\left(\tilde{\underline{\xi}}_{34} - \underline{\xi}_{34}\right)^T \tilde{\underline{\Sigma}}_{34}^{-1} \left(\tilde{\underline{\xi}}_{34} - \underline{\xi}_{34}\right) \leq \chi^2_{2,1-\alpha}, \tag{3.5.10}$$

where

$$\tilde{\underline{\xi}}_{34} = \left[\bar{b}_2 - \frac{1}{n}\left(-\frac{\bar{b}_3}{\bar{R}} - \frac{\bar{b}_2}{\bar{R}^2} + \frac{2\bar{a}_2\bar{b}_2}{\bar{R}^4}\right), \bar{a}_2 - \frac{1}{n}\left\{1 - \frac{\bar{a}_3}{\bar{R}} - \frac{\bar{a}_2(1-\bar{a}_2)+\bar{b}_2^2}{\bar{R}^2}\right\}\right]^T \tag{3.5.11}$$

and $\tilde{\underline{\Sigma}}_{34} = (\tilde{\sigma}_{ij})$ with $\tilde{\sigma}_{33} = \widetilde{\text{var}}(\bar{b}_2)$, $\tilde{\sigma}_{44} = \widetilde{\text{var}}(\bar{a}_2)$ and

$$\tilde{\sigma}_{34} = \widetilde{\text{cov}}(\bar{b}_2, \bar{a}_2) = \frac{1}{n}\left\{(1-\bar{a}_2)\bar{b}_2 + \frac{\bar{b}_4}{2} - \frac{\bar{a}_2\bar{b}_3 + \bar{b}_2\bar{a}_3}{\bar{R}} - \frac{2\bar{a}_2\bar{b}_2(1-\bar{a}_2)}{\bar{R}^2}\right\}.$$

An analogous, non-bias-corrected, confidence region follows similarly on omitting the two bias terms from (3.5.11). For a parent population which is assumed to be symmetric, $\bar{\beta}_2$ is identically zero and then a separate confidence interval for $\bar{\alpha}_2$, calculated as described above, might be considered of interest.

3.5.5 Other Joint Confidence Sets

Confidence regions for the other four possible pairs of measures can be obtained using the analogous results to those employed in the constructions described in Sections 3.5.3 and 3.5.4. Those for triples and the full set of the four measures follow by the usual extension to the results summarized in (3.5.6) and (3.5.10). As in the constructions given above, we can either include, or omit, bias-correction. Also, again as in those same constructions, the assumption of symmetry can be allowed for by equating the relevant central sine moments to zero.

3.6 A Cautionary Note

As in any situation involving the unrestricted estimation of measures defined on some finite interval, experience shows that the use of the bias-corrected estimators of Section 3.4 and the confidence set constructions described in Section 3.5 can result in inadmissible estimates. Moreover, the constructions for the confidence sets can even fail to provide a set at all. For instance, the estimate of the bias of $\bar{b}_2$

given in Section 3.5.4 can 'explode', leading to point estimates and confidence bounds beyond the interval $[-1,1]$; the argument of the inverse sine function in (3.5.3) need not necessarily be contained in $[-1,1]$; the bounds obtained using (3.5.4) or (3.5.5) can lie outside $[0,1]$ Such problems typically arise when $\bar{\alpha}_2$ and ρ are small, conditions which apply to underlying distributions which are multimodal and cyclically symmetric. Whilst, for multimodal data, we can employ the statistics considered here as global measures, their use in the analysis of the data forming the individual modes has the potential of being yet more informative. An illustrative analysis of this type is given in Section 3.7.3.

3.7 Three Illustrative Examples

In this section we apply the inferential techniques developed in Sections 3.4 and 3.5 in the analysis of three sets of circular data recorded during animal orientation experiments. The first two involve the initial directions taken by chinese painted quail on exiting two differently structured corridors. The third concerns the orientations of dragonflies with respect to the sun's azimuth.

3.7.1 Initial Headings of Chinese Painted Quail: Experiment A

The data given in Table 3.1 were collected during an orientation experiment involving chinese painted quail (*Excalfactoria chinensis*) reported in Merkel and Fischer-Klein (1973). A total of 100 birds were forced to pass, one by one, through a corridor with a change in orientation of 15° at a distance of 0.5m from its exit. On leaving the corridor the birds generally compensated for this change in orientation, their initial headings being distributed about an angle of approximately -15° from the zero direction defined by the last 0.5m of the corridor. However, as can be seen from Figure 3.1, the distribution of these angles of compensation does not appear to be symmetric.

The value of $\bar{b}_2$ for these data is –0.0216, and the bias-corrected estimate of $\bar{\beta}_2$, calculated using (3.4.1), is –0.0211. An approximate 95% confidence interval for $\bar{\beta}_2$, calculated using (3.5.8), is $(-0.0388,-0.0055)$. If the estimate of the bias is dropped from (3.5.8) the equivalent interval becomes $(-0.0383,-0.0049)$. Given that neither interval contains zero, it would appear reasonable to conclude that the parent population is asymmetric.

Table 3.1 The ordered initial headings (in degrees) of 100 chinese painted quail on exit from a dog-leg corridor.

-43	-41	-41	-37	-37	-36	-35	-35	-33	-33	-30	-30	-30	-28	-28	-28	-28	-27	-26	-26
-26	-25	-25	-25	-25	-25	-25	-25	-25	-23	-23	-23	-22	-22	-22	-22	-20	-20	-20	-20	-20
-20	-20	-18	-18	-17	-17	-17	-16	-15	-15	-15	-15	-15	-15	-15	-15	-15	-15	-15	-14	-14
-14	-13	-13	-13	-13	-13	-12	-12	-10	-10	-10	-10	-10	-9	-7	-6	-5	-5	-5	-4	-3
-3	0	0	5	6	11	12	12	13	17	17	18	18	20	40	44					

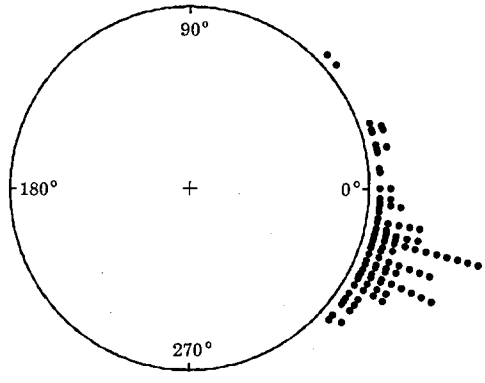

Figure 3.1 Raw circular plot of the initial headings of the 100 chinese painted quail on exit from a dog-leg corridor. The zero direction corresponds to the orientation of the last 0.5m section of the corridor.

Table 3.2 provides a summary of the point estimates and approximate 95% confidence intervals for the other three measures, calculated assuming the underlying distribution to be asymmetric. As the results obtained either using, or not using, bias-correction are identical to (at least) the quoted accuracy of two decimal places, we have only included the one set of results.

We note that the interval for μ contains the angle $-15°$, corresponding to an angle of compensation equal to minus that of the change in orientation of the last stretch of the corridor. Thus, a null hypothesis stating that the population mean angle of compensation was $\mu = -15^0$ would not be rejected in a two-sided test of size 5%.

Table 3.2 Point estimates and individual approximate 95% confidence intervals for μ, ρ and $\overline{\alpha}_2$, quoted to two decimal places. The results are those obtained when the parent population is assumed to be asymmetric.

Measure	Point estimate	Approx. 95% CI
μ	−0.26 (−14.9°)	(−0.31,−0.21) (−18.0°,−11.9°)
ρ	0.96	(0.95, 0.98)
$\overline{\alpha}_2$	0.86	(0.81, 0.91)

As has been established previously, it is unreasonable to assume the parent population to be symmetric, and hence there is little to recommend the comparison of the point estimate and confidence interval for $\overline{\alpha}_2$ with the value of the same measure for an assumed underlying wrapped normal distribution.

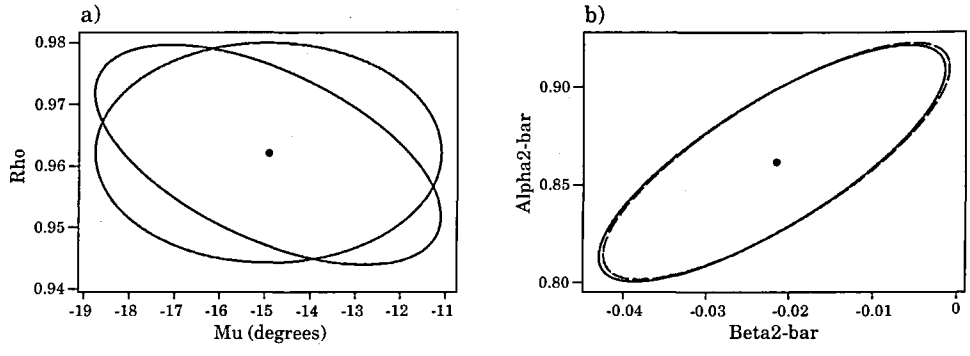

Figure 3.2 Approximate 95% confidence regions for the chinese painted quail data of Table 3.1: a) regions for (μ,ρ); b) regions for $(\overline{\beta}_2,\overline{\alpha}_2)$. In a), the inclined ellipse delimits the bias-corrected region; the other, the region without bias-correction and assuming the parent population to be symmetric. The dot represents $(\overline{\theta},\overline{R})$. In b), the solid ellipse delimits the bias-corrected region; the other, the region without bias-correction. The dot represents $(\overline{b}_2,\overline{\alpha}_2)$.

Approximate 95% confidence regions for (μ,ρ) and $(\overline{\beta}_2,\overline{\alpha}_2)$ are portrayed in Figure 3.2. In Figure 3.2a we present the bias-corrected region calculated using (3.5.6). The region obtained by dropping the bias terms from (3.5.7) is very similar and has been omitted so as not to obscure the plot. For comparative purposes we have also included the confidence region which results when the bias terms in (3.5.7) are dropped and the underlying distribution is assumed to be symmetric. Given our previous findings regarding $\overline{\beta}_2$, the use of this latter confidence region

would appear inappropriate. The bias-corrected and non-bias-corrected confidence regions for $(\bar{\beta}_2, \bar{\alpha}_2)$, calculated using (3.5.10), are represented in Figure 3.2b. We observe that the two regions are almost identical.

3.7.2 Initial Headings of Chinese Painted Quail: Experiment B

The data in Table 3.2 were recorded by Merkel and Fischer-Klein (1973) during a similar orientation experiment to that described in Section 3.7.1. In this second version of the experiment, 230 chinese painted quail were forced to pass through a completely straight, 1m long, corridor. The data correspond to the initial headings of the birds on exiting the corridor.

Table 3.3 The ordered initial headings (in degrees) of 230 chinese painted quail on exit from a straight corridor.

-44	-40	-39	-39	-37	-36	-34	-33	-31	-29	-27	-25	-23	-23	-22	-22	-20	-20	-20	-20	-18
-18	-16	-15	-15	-15	-14	-14	-14	-13	-13	-12	-12	-12	-12	-10	-10	-10	-10	-9	-9	-9
-9	-9	-9	-8	-8	-8	-7	-7	-7	-7	-7	-6	-6	-6	-6	-6	-5	-5	-5	-5	-5
-5	-5	-5	-5	-5	-4	-4	-4	-4	-4	-3	-3	-3	-3	-3	-3	-2	-2	-2	-2	-2
-2	-2	-2	-2	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
2	2	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5
5	5	5	6	6	6	6	6	7	7	7	7	7	7	7	7	7	7	8	8	8
8	9	9	9	9	9	9	10	10	10	10	10	10	10	11	11	11	11	11	12	13
13	13	13	13	14	14	14	14	15	15	15	15	16	16	17	17	17	17	18	18	19
21	21	21	22	25	26	27	30	30	31	32	33	33	33	35	35	36	44	44	51	

As can be seen from the raw circular plot of the data in Figure 3.3, the distribution of the initial headings of the quail appears to be symmetric about the zero direction defined by the orientation of the corridor. The values of $\bar{b}_2$ and $\hat{\bar{\beta}}_{2,BC}$ for these data are identical to four decimal places, both equalling 0.0005. The limits of the approximate 95% confidence intervals for $\bar{\beta}_2$, calculated using (3.5.8) with and without bias-correction, respectively, are identical to four decimal places, corresponding to an interval of $(-0.0063, +0.0073)$. Given that this interval contains the value zero, in this case it would appear reasonable to assume the parent population to be symmetric.

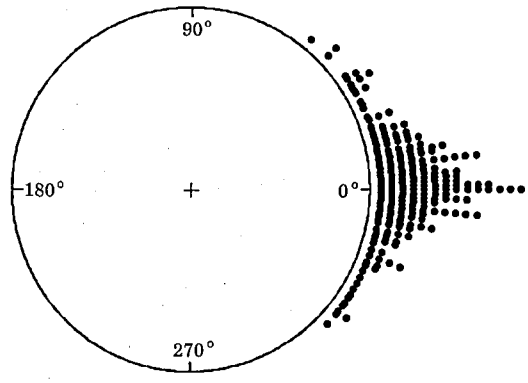

Figure 3.3 Raw circular plot of the initial headings of the 230 chinese painted quail on exit from a straight, 1m long, corridor. The zero direction corresponds to the orientation of the corridor.

In Table 3.4 we present the point estimates and approximate 95% confidence intervals for the other three measures, calculated assuming underlying symmetry. As for Table 3.2, the results obtained either using, or not using, bias-correction are identical to (at least) two decimal places, and so we have only included the one set of results.

Table 3.4 Point estimates and individual approximate 95% confidence intervals for μ, ρ and $\overline{\alpha}_2$, quoted to two decimal places. The results are those obtained when the parent population is assumed to be symmetric.

Measure	Point estimate	Approx. 95% CI
μ	0.02 (1.3°)	(−0.01, 0.06) (−0.6°, 3.3°)
ρ	0.97	(0.96, 0.97)
$\overline{\alpha}_2$	0.87	(0.85, 0.90)

Considering the confidence interval for μ, we note that it contains the angle $0°$. Hence, a null hypothesis which stated that the population mean direction corresponded to the zero direction defined by the orientation of the corridor, would not be rejected in a two-sided test of size 5%.

To three decimal places the values of $\overline{a}_2$ and $\overline{R}$ are 0.874 and 0.966, respectively. For a wrapped normal parent population we know that $\overline{\alpha}_2 = \rho^4$.

Estimating ρ by $\overline{R}$, we obtain $\hat{\overline{\alpha}}_2 = \overline{R}^4 = 0.871$ for an assumed underlying wrapped normal distribution. As the difference between $\overline{a}_2$ and $\hat{\overline{\alpha}}_2$ is only very slight, and the value 0.871 is contained in the approximate 95% confidence interval for $\overline{\alpha}_2$, one might expect the wrapped normal distribution to provide a reasonable model for these data.

In Figure 3.4 we present the approximate 95% confidence region for (μ, ρ) obtained using (3.5.6) without applying bias-correction and assuming the underlying distribution to be symmetric. If the parent population is not assumed to be symmetric the resulting bias-corrected and non-bias-corrected regions are almost identical to the region portrayed.

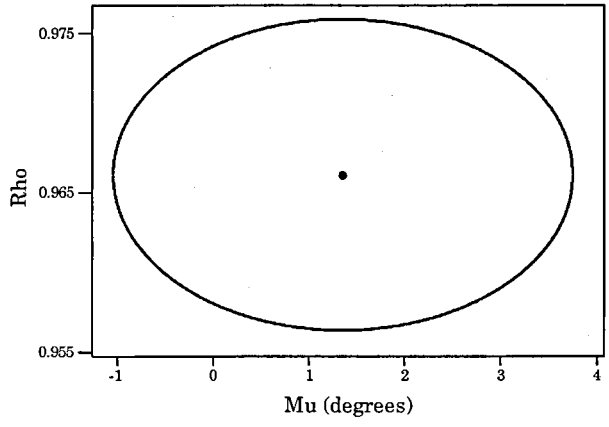

Figure 3.4 Approximate 95% confidence region for (μ,ρ) for the chinese painted quail data of Table 3.3. The ellipse delimits the non-bias-corrected region obtained assuming the parent population to be symmetric. The dot represents $(\overline{\theta},\overline{R})$.

Although we have analyzed the data in this and the preceding example as circular data, given that in both cases the data span just a reduced arc of the unit circle, we could have chosen to analyze them as linear data after applying a suitable rotation and unwrapping them to the line. In analyses analogous to the ones presented, the estimation of linear measures such as the mean, standard deviation and coefficient of skewness would then be of interest. If inference for such measures was to be based on the method of moments then the results presented in Theorem 1.1 of Chapter 1 would be relevant. The following example, however, is one which is not amenable to linear analysis, at least if the data set in its entirety is to be analyzed as a single sample.

3.7.3 Orientations of Dragonflies

As well as providing an example of a truly circular data set, i.e. one for which the data are distributed over a sizeable arc of the unit circle, our last example of this section also illustrates the importance of the cautionary note made in Section 3.6. The data set upon which our analysis is based is taken from a study reported by Hisada (1972) into the orientation of dragonflies (*Sympetrum*) with respect to the sun's azimuth. The data are reproduced in Table 3.5.

Table 3.5 The ordered orientations relative to the sun's azimuth (in degrees) of 214 dragonflies.

-127	-123	-123	-120	-116	-114	-112	-112	-112	-112	-110	-108	-108	-108
-106	-106	-106	-106	-106	-104	-104	-104	-102	-100	-100	-98	-98	-96
-96	-96	-96	-96	-94	-94	-94	-94	-94	-92	-92	-92	-92	-92
-92	-90	-90	-90	-88	-88	-88	-88	-88	-86	-86	-86	-84	-84
-84	-84	-84	-84	-82	-82	-82	-82	-82	-80	-80	-80	-80	-78
-78	-78	-78	-78	-76	-76	-76	-76	-74	-74	-74	-72	-72	-72
-70	-70	-70	-70	-67	-67	-67	-67	-63	-60	-51	-51	-47	-45
-40	-27	-27	-9	-9	-6	3	22	39	44	50	50	52	52
52	54	54	54	54	54	56	56	56	56	58	58	58	60
60	62	62	62	64	64	66	66	66	66	66	66	70	70
72	72	72	72	74	74	74	76	76	76	76	76	78	78
78	78	80	80	80	80	82	82	84	84	84	86	86	86
86	88	88	88	88	88	90	90	90	90	90	90	92	92
92	92	94	94	94	96	96	96	96	96	96	98	98	98
98	98	98	100	100	100	105	108	111	114	117	120	128	131
134	138	142	165										

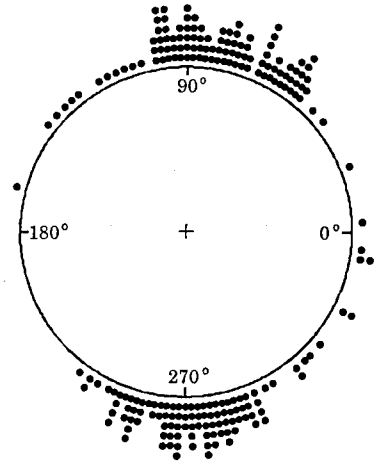

Figure 3.5 Raw circular plot of the orientations of the 214 dragonflies relative to the sun's azimuth.

From the raw circular plot in Figure 3.5 we see that the distribution of the 214 angles is clearly bimodal, with the dragonflies preferring to orientate themselves at an angle of around 90° to the left or right of the zero direction defined by the azimuth. If we were to assume the underlying distribution to be cyclically symmetric then it would be appropriate to treat the data as being axial. Here we make no such assumption and conduct inference for the basic measures of location, concentration, skewness and kurtosis corresponding to the angles as given.

Table 3.6 Non-bias-corrected (NBC) and bias-corrected (BC) point estimates for μ, ρ, $\overline{\beta}_2$, and $\overline{\alpha}_2$ quoted to two decimal places. The two bias-corrected estimates for $\overline{\alpha}_2$ are those which result from assuming asymmetry (left) and symmetry (right).

Measure	Point estimate	
	NBC	BC
μ	0.15	0.20
ρ	0.12	0.10
$\overline{\beta}_2$	0.30	10.04
$\overline{\alpha}_2$	−0.66	−1.00 −1.03

Table 3.7 Approximate 95%, non-bias-corrected (NBC) and bias-corrected (BC), confidence intervals for μ (radians), ρ, $\overline{\beta}_2$, and $\overline{\alpha}_2$ quoted to two decimal places. For both cases, the two intervals for $\overline{\alpha}_2$ are those which result from assuming asymmetry (left) and symmetry (right).

Measure	Approximate 95% confidence interval	
	NBC	BC
μ	(−0.88, 1.19)	(−0.83, 1.24)
ρ	(0.06, 0.17)	(0.05, 0.15)
$\overline{\beta}_2$	(−1.08, 1.68)	(8.66, 11.42)
$\overline{\alpha}_2$	(−1.29, −0.02) (−0.71, −0.60)	(−1.64, −0.36) (−1.09, −0.97)

In Tables 3.6 and 3.7 we present point estimates and approximate 95% confidence intervals for μ, ρ, $\overline{\beta}_2$, and $\overline{\alpha}_2$, calculated with (without) bias-correction and assuming (not assuming) symmetry. From Table 3.6 it can be seen that, whilst the estimates of μ and ρ are reasonably consistent, bias-correction has led to inadmissible estimates for $\overline{\beta}_2$ and $\overline{\alpha}_2$. Moreover, from Table 3.7 we observe that,

apart from those for ρ, the confidence intervals provide very imprecise information concerning the various measures. Despite the fact that the sample consists of 214 observations, the non-bias-corrected confidence interval for μ has a vast range, being equivalent to that of the interval (–50.42°, 68.18°). In addition, the Fisher-Lewis interval specified in (3.5.3) could not be calculated as the argument of the inverse sine function used in its construction lay outside $[-1, 1]$. We note that the limits of both intervals for $\bar{\beta}_2$ are inadmissible. Also, the non-bias-corrected interval spans the complete admissible range $[-1, 1]$, whereas its bias-corrected counterpart contains no admissible values at all. The confidence intervals for $\bar{\alpha}_2$ suffer from similar problems, with all but the non-bias-corrected interval based on the assumption of symmetry having inadmissible limits. Clearly the estimation process has broken down for these data, a consequence of the data being almost antipodally symmetric.

If instead of considering the data set in its entirety we split it in two then the problems of estimation detailed above disappear. Tables 3.8 and 3.9 contain point estimates and approximate 95% confidence intervals for the two subsamples formed from the angles falling in the two intervals $[0, \pi)$ and $[-\pi, 0)$, i.e. the 110 angles to the "left" (Sample L) and the 104 to the "right" (Sample R) of the azimuth. As the results obtained using, or not using bias-correction, and assuming symmetry, or asymmetry, differed by at most 0.01, we present just the results arising from assuming asymmetry and not applying bias correction. To two decimal places, the limits of the Fisher-Lewis interval for μ were also identical to those quoted in Table 3.9.

Table 3.8 Point estimates of μ, ρ, $\bar{\beta}_2$, and $\bar{\alpha}_2$, quoted to two decimal places, for the two subsamples formed from the angles in Table 3.5 falling to the "left" (Sample L) and "right" (Sample R) of the sun's azimuth.

Measure	Point estimate	
	Sample L	Sample R
μ	1.41 (80.8°)	–1.48 (–85.0°)
ρ	0.92	0.92
$\bar{\beta}_2$	–0.02	–0.05
$\bar{\alpha}_2$	0.73	0.75

Table 3.9 Approximate 95% confidence intervals for μ, ρ, $\overline{\beta}_2$, and $\overline{\alpha}_2$, quoted to two decimal places, for the two subsamples formed from the angles in Table 3.5 falling to the "left" (Sample L) and "right" (Sample R) of the sun's azimuth.

Measure	Approximate 95% confidence interval	
	Sample L	Sample R
μ	(1.33, 1.48) (76.2°, 84.8°)	(−1.55, −1.41) (−89.0°, −81.0°)
ρ	(0.89, 0.94)	(0.90, 0.95)
$\overline{\beta}_2$	(−0.05, 0.02)	(−0.09, −0.01)
$\overline{\alpha}_2$	(0.65, 0.80)	(0.67, 0.83)

From a consideration of the confidence intervals for $\overline{\beta}_2$ given in Table 3.9, it would appear that the parent population for Sample R is asymmetric, whereas that underlying Sample L can reasonably be assumed to be symmetric. Interestingly, the two intervals for μ do not contain the angles $90°$ and $-90°$, respectively. Indeed, given the angles they do contain, these two intervals provide evidence of a slight preference of the dragonflies to orientate themselves "towards" rather than "away" from the sun's azimuth. We conclude our analysis by noting that, given our observations regarding the estimates of $\overline{\beta}_2$ and μ, treating the data as though they were axial would seem to be unjustified.

3.8 Summary and Directions for Future Research

We conclude the present chapter with a brief summary of the major issues addressed within it, and an outline of some potential lines of related future research.

3.8.1 Summary

In this chapter we have considered the joint distribution of four fundamental statistics employed in the analysis of circular data: $\overline{\theta}$, $\overline{R}$, $\overline{b}_2$, and $\overline{a}_2$. The asymptotic joint distribution of these statistics was derived using Taylor expansion in Section 3.3. Bias-corrected estimators for the analogous population measures, μ, ρ, $\overline{\beta}_2$, and $\overline{\alpha}_2$, follow directly from the results for the asymptotic joint distribution, and were given in Section 3.4. In Section 3.5 we used those same results to derive new constructions for large-sample confidence sets for the different possible combinations of the population measures.

Problems which might be encountered when applying the derived inferential tools were identified in Section 3.6. The use of those tools, and some of the potential problems associated with their application, were illustrated in Section 3.7 using three data sets taken from the animal orientation experimentation literature.

3.8.2 Directions for Future Research

The asymptotic results derived in this chapter for the joint distribution of $\bar{\theta}$, $\bar{R}$, $\bar{b}_2$, and $\bar{a}_2$ are of wide relevance as they are nonparametric. In obtaining them, the only assumption we made regarding the underlying distribution was that $\rho \in (0, 1)$. In Section 6.4.3.2 of Chapter 6 we explore the validity of the theoretical asymptotic bias and variance results given in Theorem 3.1 when the parent population is wrapped skew-normal.

The bias-corrected estimates of the population analogues of the four statistics, and the large-sample confidence set constructions given in Sections 3.4 and 3.5, also followed directly from our asymptotic results. For the illustrative examples of Section 3.7 we found that, except when the estimation process failed completely, the results obtained with or without bias-correction, and assuming or not assuming symmetry, seldom differed greatly. However, the sizes of the samples used in those examples were fairly large. It would be interesting to compare the efficiencies of the different estimators for data sets of varying size sampled from different forms of circular distributions. This idea is pursued in Section 6.4.3.3 of Chapter 6 for the case of a parent population assumed to be wrapped skew-normal. Similarly, a comparison of the actual coverages of the proposed large-sample confidence set constructions also suggests itself as being most valuable. Related to this last point, in Section 4.4 of Chapter 4 we present results for the operating characteristics of a test for reflective symmetry based on the statistic $\bar{b}_2$.

Regarding theoretical results associated with the above ideas, we note that Mardia (1972, Sections 4.5.3 and 4.9.3) considers the joint distribution of $\bar{\theta}$ and $\bar{R}$ for an assumed von Mises parent population. For the same underlying distribution, Fisher & Lewis (1983) compare the form of the confidence interval for μ obtained using their construction with that of the corresponding exact interval given by Mardia (1972, pp. 145-6). Extending these ideas, it might be possible to obtain theoretical results for the joint distributions of other combinations of the four statistics for the von Mises distribution, or, indeed, any other appealing form of

parent population. Those joint distributions might then be used to obtain distribution-specific estimators and confidence set constructions for the population measures μ, ρ, $\overline{\beta}_2$, and $\overline{\alpha}_2$. However, on the face of it, obtaining theoretical results of this type would appear to be somewhat of a tall order, at least for the full joint distribution of all four statistics. Moreover, for a specific parent population, the measures μ, ρ, $\overline{\beta}_2$, and $\overline{\alpha}_2$ may not be of primary interest. For instance, for an underlying wrapped normal distribution, the parameter σ rather than the measure ρ would be of interest, and $\overline{\beta}_2$ and $\overline{\alpha}_2$ would be of no real interest at all. Nevertheless, for skew distributions we might envisage all four measures to be relevant, and theoretical results such as those described above would be most valuable.

Our results are based on asymptotic theory and therefore should not be applied in the analysis of small sized samples. For situations in which our asymptotic results are thought not to apply we can always use computer intensive methods of inference. Fisher & Hall (1989, 1990) and Fisher (1993, Chapter 8) consider bootstrap procedures for obtaining confidence intervals for μ. The ideas underlying those procedures could be adapted to obtain computer intensive methods of inference for the other three measures ρ, $\overline{\beta}_2$, and $\overline{\alpha}_2$. For the construction of bootstrap confidence sets for pairs, triples and the full set of the four measures, the approach of Yeh & Singh (1997) referred to in Section 2.8.2 of Chapter 2 suggests itself as being potentially useful.

Chapter 4 Testing for Circular Reflective Symmetry About an Unknown Central Direction

4.1 Introduction

Symmetry is one of the most basic of dividing hypotheses, the acceptance of which leads to the subsequent exploration of models which, when compared with their skew competitors, generally have relatively concise parametric specifications. On the other hand, the rejection of symmetry raises important issues as to precisely which of a distribution's characteristics are of real statistical interest. In this chapter we consider in detail the testing of circular data for a particular form of symmetry, namely, reflective symmetry about an unknown central direction, against asymmetry.

In Section 4.2 we provide a brief review of the enormous wealth of procedures which have been proposed for testing linear data for reflective symmetry about an unknown centre. In addition, we describe the various forms of symmetry associated with circular data which might be tested for, and cite the only published work that we are aware of regarding the testing of circular data for reflective symmetry about an unknown central direction.

In Section 4.3 we introduce a new large-sample omnibus procedure for testing circular data for reflective symmetry about an unknown central direction, the test and its asymptotic sampling properties following directly from the results given in Theorem 3.1 of Chapter 3. The operating characteristics of the test are explored in Section 4.4, the asymptotic power of the test being given in Section 4.4.1. In Section 4.4.2 we describe a Monte Carlo experiment designed to investigate the small-sample operating characteristics of the test. In this study we simulate data from a variety of unimodal distributions, three of which are new to the literature. The basic properties of these three distributions are given in Sections 4.4.2.1.1, 4.4.2.1.2 and 4.4.2.2.1. Results for the test's ability to maintain the nominal significance level for underlying symmetric distributions, and its power against skew alternatives, are presented in Section 4.4.2.4.

Computer intensive versions of the new test are outlined in Section 4.5. We describe bootstrap and randomization variants of the test based on the use of the symmetrization device of Efron (1979). The results from a small scale simulation study conducted in order to compare the operating characteristics of the large-sample, bootstrap and randomization versions of the test are presented in Section 4.6.

In Section 4.7 we illustrate the application of the developed testing methodology with analyses of seven data sets taken from the circular statistics literature. The chapter closes, in Section 4.8, with a summary of its content and an indication of potential future lines of research related to the testing of circular data for reflective symmetry about an unknown central direction.

In the main, the content of Sections 4.1-4.7 draws heavily on that of Pewsey (2002b). However, large parts of Sections 4.4 and 4.7 are new, and the detail given in Section 4.6 is greater than that presented in the cited paper.

4.2 Background to Testing for Symmetry on the Line and Circle

Numerous procedures have been proposed in the literature for testing univariate data on the line for reflective symmetry. These tests divide into two main groups, the division resulting from whether the centre of symmetry is specified (or assumed known) or not. The former group of tests is considered in detail in Chapter 5. The second group of procedures, designed for testing the non-parametric hypothesis of reflective symmetry about an unknown centre, against asymmetry, is that which is of direct relevance to our deliberations in the present chapter. Amongst the multifarious tests in this group we cite those of: Gupta (1967), based on the coefficient of skewness; Gupta (1967), Antille *et al.* (1982) and Bhattacharya *et al.* (1982), involving modifications of the one-sample Wilcoxon test centred on the sample median; Gastwirth (1971) and Cabilio & Masaro (1996), founded upon the difference between the sample mean and median; Finch (1977) and Antille *et al.* (1982), involving the size of the gaps between observations; Davis & Quade (1978) and Randles *et al.* (1980), based on the skewness evident in triples of observations; Boos (1982) and Koziol (1983), incorporating Cramér-von Mises statistics; Csörgö & Heathcote (1987), based on the empirical characteristic symmetry function; Schuster & Barker (1987), involving bootstrapping samples from the symmetric distribution closest to the empirical distribution function.

The development of methods for testing circular data for symmetry has been nowhere near as extensive. In addition, the situation on the circle is somewhat more involved as, due to its compactness and the isometries of rotation and reflection, "symmetry" is not uniquely defined. Consequently, there are three, rather than two, types of procedure associated with symmetry. The first, that for testing for l-fold, or cyclic, symmetry, has no equivalent on the line. Jupp & Spurr (1983) introduced rank based procedures which can be used to test for this form of symmetry. The second type of procedure is designed for testing for reflective symmetry about a specified or known centre. An in-depth treatment of procedures of this type is reserved until Chapter 5.

It is the third and arguably most important type of procedure, designed to test the null hypothesis of reflective symmetry about an unknown central direction, henceforth H_0, against the alternative hypothesis of an asymmetric parent population, which we consider at length in the present chapter. Despite the fact that, as Mardia (1972, p. 10) states, "symmetrical distributions on the circle are comparatively rare", the investigation of this pair of hypotheses has received virtually no attention in the literature. This oversight can perhaps best be explained by the dominating theoretical role of the von Mises distribution in the analysis of circular data and the tacit assumption of symmetry associated with its use.

Fisher (1993, Section 4.2) introduces an exploratory graphical technique for assessing whether a unimodal distribution is symmetric. He also suggests, p. 80, that a formal test of H_0 can be based on a standard test for linear data, such as the Wilcoxon signed-rank test, with the unknown centre of symmetry estimated by the sample median. However, it is inappropriate to apply such tests in this way as they are no longer non-parametric when the centre of symmetry has been estimated, the significance level and power generally being considerably lower than their nominal levels; see, for example, Gastwirth (1971).

Of course, if the data being analyzed are distributed over a reduced arc of the circle then H_0 can be tested using any one of the linear tests referred to previously, after first applying a suitable rotation to the data and unwrapping them to the line. However, if the data cover a sizeable arc of the circle then these tests cannot be employed as the statistical summaries and distribution theory upon which they are based no longer apply. Thus there is a genuine need for procedures which can be

employed to test truly circular data, covering a substantial arc of the circle, for reflective symmetry about an unknown central direction.

4.3 A Large-sample Omnibus Test for Circular Reflective Symmetry About an Unknown Central Direction

Our omnibus test of H_0 is based on the asymptotic results for the sample second sine moment about the mean direction, $\bar{b}_2$, given in Theorem 3.1 of Chapter 3. We note that the statistic $\bar{b}_2$ had previously been proposed by Cox (1975) as the basis of a test for skew departures from the von Mises distribution.

For a distribution which is reflectively symmetric about a central direction ϕ we have $f(\theta) = f(2\phi - \theta)$ and, clearly, $f(\theta)$ is also symmetric about $\phi + \pi$. If, in addition, $\rho \neq 0$ then $\mu = \phi$ and $E(\bar{b}_2) = 0$. An estimate of the variance of $\bar{b}_2$ in this context can be obtained using the expression for $\text{var}(\bar{b}_2)$ in (3.3.2) with $\bar{\beta}_2 = 0$ and ρ and $\bar{\alpha}_p$ replaced by their plug-in estimates, $\bar{R}$ and $\bar{a}_p = \frac{1}{n}\sum_{i=1}^{n} \cos p(\theta_i - \bar{\theta})$, respectively. Thus an asymptotically distribution-free test of H_0 can be based on the studentized statistic

$$\bar{b}_2 \Big/ \{\hat{\text{var}}(\bar{b}_2)\}^{1/2}, \tag{4.3.1}$$

where

$$\hat{\text{var}}(\bar{b}_2) = \frac{1}{n}\left[\frac{1-\bar{a}_4}{2} - 2\bar{a}_2 + \frac{2\bar{a}_2}{\bar{R}}\left\{\bar{a}_3 + \frac{\bar{a}_2(1-\bar{a}_2)}{\bar{R}}\right\}\right],$$

which is implicitly assumed to be positive. Large absolute values of (4.3.1) compared with the quantiles of the standard normal distribution lead to the rejection of symmetry in favour of some skewed alternative.

Given the assumption of Theorem 3.1 that $\rho \in (0, 1)$, our test can be used with data from most, but not all, underlying distributions. The condition that $\rho \neq 0$ excludes the uniform distribution and certain multimodal distributions. The latter include all multimodal distributions which are either cyclically symmetric, or have more than one axis of reflective symmetry, or both. However, if it is thought possible that $\rho = 0$ for some underlying multimodal distribution, and the number of modes, *m*, of the distribution can be established beforehand, then the test can be

applied after first employing the device of m-fold wrapping of the circle onto itself (see Mardia & Jupp 1999, p. 53). This device consists of multiplying each data value by m, by so doing producing a sample from a distribution with a uniquely defined central direction. In fact this procedure can be used with data from any form of multimodal distribution, and so if there is any doubt as to whether $\rho \neq 0$ it would be advisable to apply the test to both the original data and to the wrapped data and compare results. Turning our attention to the uniform distribution, it is the unique distribution which is invariant under both isometries of reflection and rotation. Isotropy plays such a central role in the analysis of circular data that numerous methods are available for testing for it (see Mardia & Jupp 1999, Chapter 6). The other extreme, when $\rho = 1$, corresponds to the pathological case of a point distribution.

4.4 The Operating Characteristics of the Test

When assessing the performance of a test of symmetry there are two general requirements which it should meet. Firstly, it should be able to maintain the nominal significance level over a wide range of symmetrical distributions. Secondly, it should be powerful against asymmetric alternatives. In Section 4.4.1 we consider the asymptotic power of the test. Results for the small-sample operating characteristics of the test are presented in Section 4.4.2.

4.4.1 The Asymptotic Power of the Test

Using the results of Theorem 3.1, under H_0 the asymptotic distribution of $\bar{b}_2$ is normal with $E_0(\bar{b}_2) = 0$ and

$$\mathrm{var}_0(\bar{b}_2) = \frac{1}{n}\left[\frac{1-\bar{\alpha}_4}{2} - 2\bar{\alpha}_2 + \frac{2\bar{\alpha}_2}{\rho}\left\{\bar{\alpha}_3 + \frac{\bar{\alpha}_2(1-\bar{\alpha}_2)}{\rho}\right\}\right],$$

assuming $\mathrm{var}_0(\bar{b}_2)$ to be positive. If the parent population is asymmetric then the asymptotic distribution of $\bar{b}_2$ is again normal but with

$$E_1(\bar{b}_2) = \bar{\beta}_2 + \frac{1}{n}\left(-\frac{\bar{\beta}_3}{\rho} - \frac{\bar{\beta}_2}{\rho^2} + \frac{2\bar{\alpha}_2\bar{\beta}_2}{\rho^4}\right)$$

and

$$\mathrm{var}_1\left(\bar{b}_2\right)=\frac{1}{n}\left[\frac{1-\bar{\alpha}_4}{2}-2\bar{\alpha}_2-\bar{\beta}_2^2+\frac{2\bar{\alpha}_2}{\rho}\left\{\bar{\alpha}_3+\frac{\bar{\alpha}_2\left(1-\bar{\alpha}_2\right)}{\rho}\right\}\right],$$

assuming $\mathrm{var}_1\left(\bar{b}_2\right)$ to be positive. It therefore follows that for a significance level of $100\alpha\,\%$ the asymptotic power of the test is

$$1-\Phi\left[\frac{z_{\alpha/2}\left\{\mathrm{var}_0\left(\bar{b}_2\right)\right\}^{1/2}-E_1\left(\bar{b}_2\right)}{\left\{\mathrm{var}_1\left(\bar{b}_2\right)\right\}^{1/2}}\right]+\Phi\left[\frac{-z_{\alpha/2}\left\{\mathrm{var}_0\left(\bar{b}_2\right)\right\}^{1/2}-E_1\left(\bar{b}_2\right)}{\left\{\mathrm{var}_1\left(\bar{b}_2\right)\right\}^{1/2}}\right]. \qquad (4.4.1)$$

The form taken by (4.4.1) as a function of ρ is exhibited in Section 4.4.2.4.1 for a range of skew parent populations.

4.4.2 Monte Carlo Investigation of the Small-sample Operating Characteristics of the Test

In order to assess the test's ability to maintain the nominal significance level for symmetric parent populations, and its power against asymmetric alternatives, we conducted a simulation study using a range of underlying unimodal distributions with mean direction the origin.

4.4.2.1 Symmetric Models

Four symmetric distributions were used in the study, the first three being the wrapped normal, wrapped Laplace and wrapped Cauchy distributions. On the line, the unwrapped versions of these distributions are classified as being symmetric short, medium and heavy-tailed, respectively. The fourth was a mixture distribution having wrapped normal and uniform components, the former with standard deviation set equal to 0.4590. The properties of the wrapped normal and wrapped Cauchy distributions are well known; see, for example, Mardia & Jupp (1999, pp. 50-52). As the other two distributions are new to the literature, we briefly review those of their basic properties most relevant to this and the next chapter.

4.4.2.1.1 Fundamental Properties of the Wrapped Laplace Distribution

Consider the (linear) double exponential random variable X, centred at the origin, with probability density function

$$f(x;\lambda)=\frac{\lambda}{2}\exp\left(-\lambda|x|\right) \qquad \lambda>0,-\infty<x<\infty,$$

and characteristic function

$$\psi_X(t)=\frac{\lambda^2}{\lambda^2+t^2}.$$

Then the probability density function of $\theta = X \pmod{2\pi}$ is

$$f(\theta;\lambda)=\frac{\lambda}{2}\sum_{k=-\infty}^{\infty}\exp\left(-\lambda|\theta+2\pi k|\right),$$

and its characteristic function $\{\psi_{\theta,p} : p = 0,\pm1,...\}$ is given by

$$\psi_{\theta,p}=\frac{\lambda^2}{\lambda^2+p^2}.$$

As the distribution is symmetric about 0, it follows that $\mu = \tilde{\mu} = 0$, where $\tilde{\mu}$ denotes the median direction. Also, $\beta_p = 0$, $\alpha_p = \psi_{\theta,p}$, and thus $\rho = \alpha_1 = \lambda^2/(1+\lambda^2)$. Inverting this last expression, we obtain $\lambda = \{\rho/(1-\rho)\}^{1/2}$.

4.4.2.1.2 Fundamental Properties of the Wrapped Normal and Uniform Mixture Distribution

Let θ be a circular random variable with probability density function

$$f(\theta;q,\rho_{WN})=q\,f_{WN}(\theta;\rho_{WN})+(1-q)\frac{1}{2\pi},$$

where $f_{WN}(\theta;\rho_{WN})$ denotes the density of the wrapped normal distribution with mean direction 0 and mean resultant length $\rho_{WN} = e^{-\frac{1}{2}\sigma^2}$. Then θ has a mixture distribution with wrapped normal and uniform components and mixing proportions $q \in [0,1]$ and $(1-q)$. Thus, its distribution is a special case of the symmetric wrapped stable – circular uniform mixture family considered by Sengupta & Pal (2001). When $q = 1$ the density reduces to that of the wrapped normal component, and when $q = 0$ θ is a circular uniform random variable. We can write,

$$f(\theta;q,\rho_{WN})=\frac{q}{\sqrt{2\pi}\sigma}\sum_{k=-\infty}^{\infty}\exp\left\{\frac{-(\theta+2\pi k)^2}{2\sigma^2}\right\}+\frac{(1-q)}{2\pi}.$$

Once again, as the distribution of θ is symmetric about 0, $\mu = \tilde{\mu} = 0$ and $\beta_p = 0$. Also, $\alpha_p = \psi_{\theta,p}$, with $\psi_{\theta,0} = 1$ and, for $p = \pm1,\pm2,...$,

$$\alpha_p = E(\cos p\theta) = q\int_0^{2\pi} \cos p\theta \, f_{WN}(\theta;\rho_{WN})d\theta + \frac{(1-q)}{2\pi}\int_0^{2\pi} \cos p\theta \, d\theta$$

$$= q\rho_{WN}^{p^2}.$$

Hence, the mean resultant length for θ, ρ, equals $\alpha_1 = q\rho_{WN} = qe^{-\frac{1}{2}\sigma^2}$. For the specified value used in our simulation study of $\sigma = 0.4590$, and $\rho \in \left[0, e^{-\frac{1}{2}\sigma^2}\right] \approx [0, 0.9]$, $q = \rho / e^{-\frac{1}{2}\sigma^2}$. Thus, as ρ ranges between 0 and $e^{-\frac{1}{2}\sigma^2}$, the distribution of θ varies between being that of a uniform random variable and that of a wrapped normal random variable, the latter with mean direction 0 and mean resultant length $e^{-\sigma^2/2} \approx 0.9$. Consequently, the heavy-tailedness of the distribution depends on ρ.

Linear plots of the densities of the four symmetric distributions considered in the simulation study are given in Figure 4.1. The densities shown correspond to the respective distributions for a mean resultant length of 0.45.

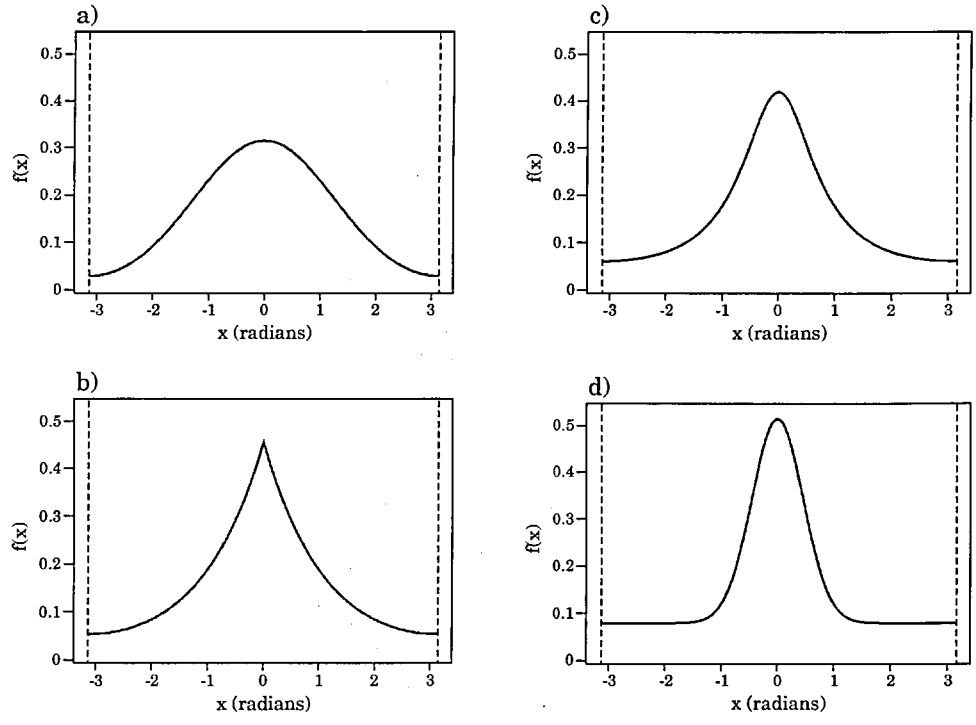

Figure 4.1 Linear plots of the symmetric densities used in the simulation study, with μ = 0 and ρ = 0.45: a) wrapped normal, b) wrapped Laplace, c) wrapped Cauchy, d) wrapped normal and uniform mixture. The dashed vertical lines delimit the bounds $-\pi$ and π.

4.4.2.2 Asymmetric Models

As asymmetric models we used the wrapped exponential distribution and three cases of the wrapped skew-normal distribution on the circle (WSNC) of Pewsey (2000b), the latter with values of the associated skewness parameter, λ, of 2, 5 and ∞. The wrapped exponential distribution is new to the literature and so in Section 4.4.2.2.1 we provide a brief summary of its basic properties. Also, in Section 4.4.2.2.2, we outline those properties of the WSNC distribution most relevant to this and the next chapter, a full treatment of the distribution being given in Chapter 6.

4.4.2.2.1 Fundamental Properties of the Wrapped Exponential Distribution

Consider the (linear) exponential random variable X, with probability density function

$$f(x;\lambda)=\lambda\exp(-\lambda x) \qquad \lambda>0, x>0,$$

and characteristic function

$$\psi_X(t)=\frac{\lambda}{\lambda-it}=\frac{\lambda^2+i\lambda t}{\lambda^2+t^2}.$$

The circular random variable $\theta = X \pmod{2\pi}$ has probability density function

$$f(\theta;\lambda)=\lambda\sum_{k=0}^{\infty}\exp\{-\lambda(\theta+2\pi k)\}$$

and characteristic function $\{\psi_{\theta,p} : p = 0,\pm1,...\}$ given by

$$\psi_{\theta,p}=\frac{\lambda^2+i\lambda p}{\lambda^2+p^2}.$$

Thus, $\alpha_p=\lambda^2/(\lambda^2+p^2)$ and $\beta_p=\lambda\, p/(\lambda^2+p^2)$. The mean direction is given by $\mu=\tan^{-1}(\beta_1/\alpha_1)=\tan^{-1}(1/\lambda)$, and the mean resultant length is $\rho=\{\alpha_1^2+\beta_1^2\}^{1/2}=\lambda/(1+\lambda^2)^{1/2}$. For a specified value of ρ the corresponding value of λ is thus $\lambda=\rho/(1-\rho^2)^{1/2}$. The median direction, $\tilde{\mu}$, is such that

$$\int_{\tilde{\mu}}^{\tilde{\mu}+\pi} f(\theta;\lambda)\,d\theta=\frac{1}{2}.$$

As

$$\begin{aligned}\int f(\theta;\lambda)\,d\theta &= \int \sum_{k=0}^{\infty} \lambda \exp\{-\lambda(\theta+2\pi k)\}\,d\theta \\ &= \sum_{k=0}^{\infty} \int \lambda \exp\{-\lambda(\theta+2\pi k)\}\,d\theta \\ &= \sum_{k=0}^{\infty} -\exp\{-\lambda(\theta+2\pi k)\} + C,\end{aligned}$$

then

$$\begin{aligned}\int_{\tilde{\mu}}^{\tilde{\mu}+\pi} f(\theta;\lambda)\,d\theta &= \sum_{k=0}^{\infty} -\exp\{-\lambda(\tilde{\mu}+\pi+2\pi k)\} - \sum_{k=0}^{\infty} -\exp\{-\lambda(\tilde{\mu}+2\pi k)\} \\ &= e^{-\lambda\tilde{\mu}}\left\{\left(1-e^{-\pi\lambda}\right)\sum_{k=0}^{\infty} e^{-2\pi\lambda k}\right\}.\end{aligned}$$

It follows that

$$\tilde{\mu} = -\frac{1}{\lambda}\log_e\left\{\frac{1-e^{-2\pi\lambda}}{2\left(1-e^{-\pi\lambda}\right)}\right\}.$$

4.4.2.2.2 Fundamental Properties of the Wrapped Skew-normal Distribution on the Circle

We show in Section 6.4 of Chapter 6 that if the $\mathrm{SN}_{\mathrm{D}}(\xi,\eta,\lambda)$ distribution is wrapped onto the circle then the mean direction and mean resultant length of the resulting wrapped skew-normal distribution on the circle are given by

$$\mu = \tan^{-1}\left[\{\sin\xi + \Im(\delta\eta)\cos\xi\}/\{\cos\xi - \Im(\delta\eta)\sin\xi\}\right] \tag{4.4.2}$$

and

$$\rho = \omega\left\{1+\Im^2(\delta\eta)\right\}^{1/2}, \tag{4.4.3}$$

where $\delta = \lambda/\left(1+\lambda^2\right)^{1/2}$, $\omega = e^{-\frac{1}{2}\eta^2}$ and for $x>0$, $\Im(x)=\int_0^x be^{\frac{1}{2}u^2}du$, $\Im(-x)=-\Im(x)$ and $b=(2/\pi)^{1/2}$. For given values of λ and ρ, one can use (4.4.3) to solve numerically for η. From (4.4.2), when $\xi=0$, the mean direction, μ_0, is given by $\mu_0 = \tan^{-1}\{\Im(\delta\eta)\}$. Thus, for a general value of ξ, $\mu = \xi + \mu_0 \pmod{2\pi}$. It follows that $\xi = -\mu_0$ when $\mu = 0$.

For specified values of λ and η, and $\xi = 0$, the median direction, $\tilde{\mu}_0$, can be obtained using numerical methods as the solution to

$$\frac{2}{\eta}\int_{\tilde{\mu}_0}^{\tilde{\mu}_0+\pi}\sum_{k=-\infty}^{\infty}\phi\left(\frac{\theta+2\pi k}{\eta}\right)\Phi\left\{\lambda\left(\frac{\theta+2\pi k}{\eta}\right)\right\}d\theta=\frac{1}{2}.$$

For a general value of ξ, the median direction, $\tilde{\mu}$, is then given by $\tilde{\mu}=\xi+\tilde{\mu}_0 \pmod{2\pi}$. It follows that when $\tilde{\mu}=0$, $\xi=-\tilde{\mu}_0$.

Linear plots of the densities of the wrapped exponential distribution and the wrapped skew-normal distribution with λ = 2, 5 and ∞ are given in Figure 4.2. The four densities correspond to the respective distributions with ρ = 0.45 and mean direction 0. As can be seen from this figure, the distribution with λ = 2 is only moderately skew, while $\lambda=\infty$ corresponds to the highly skew wrapped half-normal distribution. The wrapped exponential distribution is even more skew.

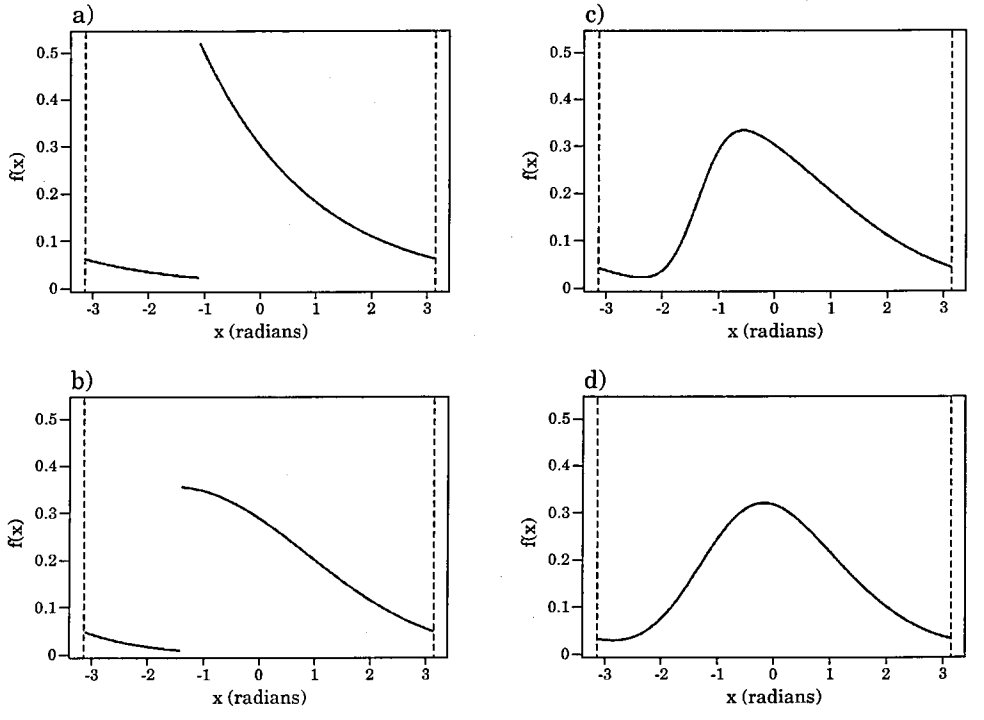

Figure 4.2 Linear plots of the asymmetric densities used in the simulation study, with μ = 0 and ρ = 0.45: a) wrapped exponential, b) wrapped half-normal, c) WSNC (λ=5), d) WSNC (λ=2). The dashed vertical lines delimit the bounds $-\pi$ and π.

4.4.2.3 Further Design Features of the Simulation Experiment

The two operating characteristics of the test were explored for sample sizes of 20, 30, 50, 100, 200 and 500, and, except for the mixture distribution, ρ-values of 0.1(0.05)0.95. For the wrapped normal and uniform mixture distribution the maximum ρ-value was set at 0.9, close to its maximum possible value (see Section 4.4.2.1.2). The empirical size and power values were calculated using 3000 simulated samples for each distribution, sample size and ρ combination. The nominal significance levels investigated were 10%, 5% and 1%.

4.4.2.4 Results

In the following two subsections we present results for the two operating characteristics of the test separately. Only the results for a nominal significance level of 5% are given, the general conclusions arising from them also being representative of those for the other two significance levels investigated in the simulation study.

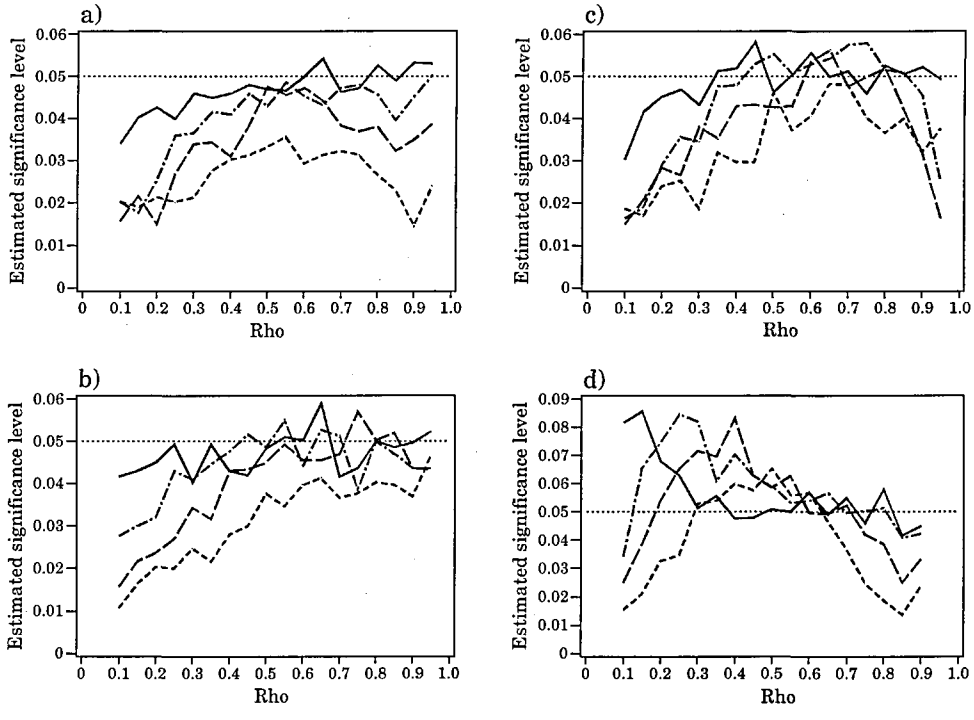

Figure 4.3 Estimated size of the test for the: a) wrapped normal, b) wrapped Laplace, c) wrapped Cauchy, d) wrapped normal and uniform mixture distributions. Sample sizes represented are: – – – – – (n = 20); — — — (n = 50); — · — (n = 100); ——— (n = 500). The dotted horizontal line delimits the nominal significance level of α = 0.05.

4.4.2.4.1 Ability to Maintain the Nominal Significance Level

Figure 4.3 displays the results for this operating characteristic of the test. We have omitted the results for sample sizes of 30 and 200 from the four plots making up this figure so as not to obscure their content. As each empirical result represented in these plots was obtained from 3000 simulated samples, the standard error of any such result is approximately 0.004. From a consideration of the plots we observe that the test is generally conservative for small sized samples. As might be expected, the test's ability to hold the nominal significance level improves with increasing sample size and concentration of the parent population. The results for the mixture distribution portrayed in Figure 4.3d are somewhat discordant with those in the other three plots, as they indicate that the test is non-conservative for certain sample size and ρ combinations. For this parent population, the test's non-conservatism appears to increase with sample size and diminishing concentration.

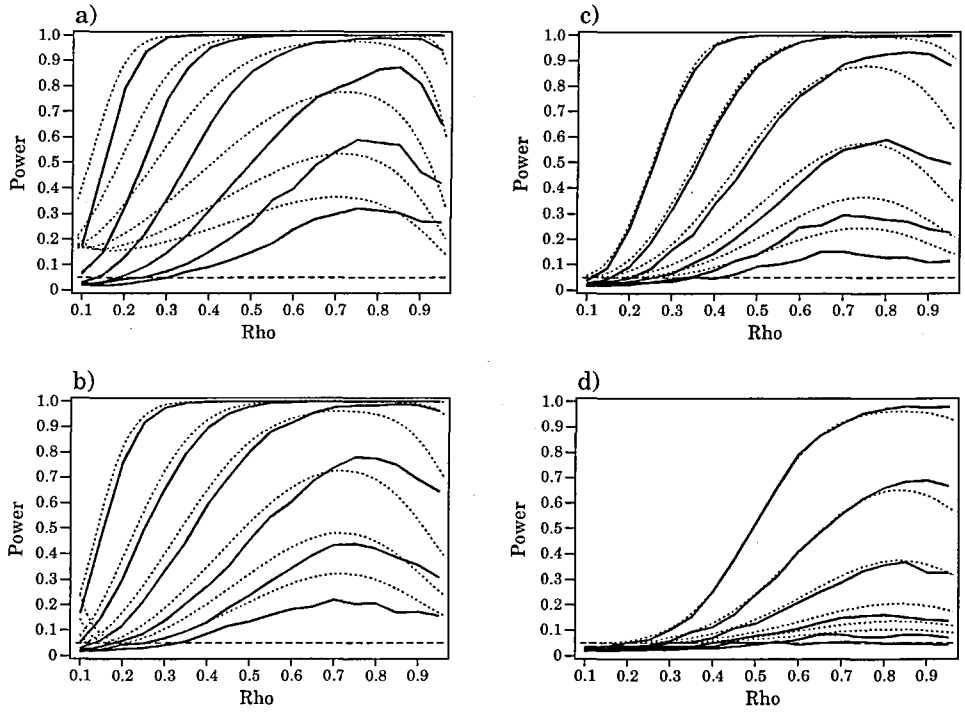

Figure 4.4 Theoretical asymptotic power (· · · · · ·) and empirical power (———) of the test for the: a) wrapped exponential, b) wrapped half-normal, c) WSNC (λ=5), d) WSNC (λ=2) distributions. The six curves of each type correspond to sample sizes of 20, 30, 50, 100, 200 and 500, the power increasing with sample size. The dashed horizontal line delimits the nominal significance level of $\alpha = 0.05$.

4.4.2.4.2 Power Against Skew Alternatives

The empirical power results for the test are displayed in Figure 4.4. For this figure, the standard error of any result represented is, at most, 0.009. Also included in the four plots making up this figure are the corresponding theoretical asymptotic power functions calculated using (4.4.1). It can be seen that there is reasonable agreement between the two sets of power curves for moderate to large sized samples from moderately skew parent populations which are not highly dispersed or highly concentrated. The disparities between the theoretical and empirical power curves are greatest for the more skew parent populations and the extremes of the concentration scale. Regarding this last point, we note that the theoretical results upon which the new test is based do not apply for the extreme values of ρ of 0 and 1. As is to be expected, the true (empirical) power of the test increases with increasing sample size. The power also increases with ρ before reaching a maximum and subsequently decreasing as ρ increases further. This deterioration in power is particularly abrupt for moderately sized samples from highly skew parent populations. Of course, $\rho = 1$ corresponds to a point distribution and for such a distribution $\bar{b}_2$ is identically zero.

4.5 Randomization and Bootstrap Versions of the Test

The device of symmetrizing a sample introduced by Efron (1979) can be used to define randomization and bootstrap versions of the test based on the pivotal statistic (4.3.1). This device produces a symmetric sample of size $2n$ by augmenting the original sample with the values $2\hat{\phi} - \theta_i$, where $\hat{\phi}$ is an estimate of the central direction. Two obvious estimates of ϕ are $\bar{\theta}$ and the sample median direction, $\tilde{\theta}$, the latter obtained by minimizing the sample circular mean deviation, $d(\varphi) = \pi - \frac{1}{n}\sum_{i=1}^{n}\left|\pi - \left|\theta_i - \varphi\right|\right|$. The former is an appropriate estimate so long as $\rho \neq 0$, it is relatively easy to compute and is the choice favoured by Fisher (1993, p. 205). Use of $\tilde{\theta}$ requires the assumption that the median direction is unique and equal to ϕ. Whilst this will be the case for a symmetric unimodal distribution, it may not be for a multimodal distribution. The precise conditions required to ensure uniqueness of the median direction are specified by Purkayastha (1995). The

robustness of $\bar{\theta}$ and $\tilde{\theta}$ has been studied by Lenth (1981), Ko and Guttorp (1988), and He & Simpson (1992). This work shows that $\tilde{\theta}$ is the more robust, although Fisher (1993, p. 72) contends that the conditions under which there is likely to be any great improvement in performance are unlikely to be encountered with real data.

Once a symmetrized sample has been produced, randomization and bootstrap samples can be generated from it and fed into their respective testing frameworks. Here a randomization sample is formed by sampling an element at random from each of the n pairs $(\theta_i, 2\hat{\phi}-\theta_i)$, whereas a bootstrap sample is generated by sampling n elements at random and with replacement from the symmetrized sample. For small samples "exact" tests can be carried out using complete enumeration of the resampling distribution. For larger samples, the variance reduction technique of balanced resampling can be employed if the storage of large arrays is not a problem. The hypothesis H_0 is rejected if the value of (4.3.1) for the original sample is judged to be extreme when compared to its resampling distribution.

4.6 A Comparative Monte Carlo Experiment

In order to compare the small-sample properties of the large-sample version of the test with those of its randomization and bootstrap counterparts, we carried out a small scale simulation experiment. The experiment was based on the same unimodal distributions as employed in the Monte Carlo study of Section 4.4.2. However, given the vast increase in computer time required to implement the randomization and bootstrap versions of the test, only three ρ-values were considered: 0.2, 0.45 and 0.7. The sample sizes explored were 10, 20, 30, 50 and 100. For each distribution, ρ and sample size combination, we simulated 3000 samples. In turn, for each one of these samples, 2000 basic bootstrap and 2000 randomization samples of the same size were generated from the associated symmetrized sample, with either $\bar{\theta}$ or $\tilde{\theta}$ being used as an estimate of ϕ. The nominal significance levels investigated were 10%, 5% and 1%.

The results obtained from the study are summarized in Tables 4.1 and 4.2. Again, we present only those results for a nominal significance level of 5%, these being representative of the results for all three significance levels considered. From

Table 4.1 Estimated significance level of the large-sample (LS), bootstrap (B) and randomization (R) versions of the proposed test as a function of the underlying symmetric distribution, ρ and n. The two options used in the symmetrization of the samples are denoted by (mean) and (median). The nominal significance level is $\alpha = 0.05$.

Distribution	Test version	n	10			20			30			50			100		
		ρ	0.20	0.45	0.70	0.20	0.45	0.70	0.20	0.45	0.70	0.20	0.45	0.70	0.20	0.45	0.70
Wrapped normal	LS		0.013	0.024	0.020	0.019	0.033	0.025	0.018	0.038	0.038	0.027	0.038	0.038	0.026	0.037	0.049
	B (mean)		0.035	0.051	0.034	0.032	0.047	0.038	0.035	0.048	0.048	0.045	0.043	0.043	0.040	0.041	0.052
	B (median)		0.029	0.028	0.013	0.030	0.052	0.043	0.032	0.050	0.048	0.038	0.053	0.046	0.035	0.055	0.055
	R (mean)		0.088	0.094	0.081	0.087	0.075	0.050	0.085	0.063	0.051	0.089	0.051	0.047	0.062	0.045	0.053
	R (median)		0.067	0.063	0.033	0.074	0.073	0.053	0.065	0.064	0.053	0.075	0.062	0.048	0.055	0.060	0.056
Wrapped Laplace	LS		0.025	0.019	0.022	0.019	0.034	0.040	0.021	0.036	0.047	0.019	0.045	0.056	0.035	0.040	0.051
	B (mean)		0.039	0.044	0.081	0.040	0.045	0.057	0.041	0.052	0.058	0.040	0.053	0.059	0.049	0.045	0.052
	B (median)		0.029	0.037	0.012	0.031	0.047	0.036	0.035	0.047	0.052	0.033	0.051	0.055	0.045	0.056	0.045
	R (mean)		0.087	0.090	0.102	0.090	0.066	0.079	0.091	0.069	0.062	0.085	0.062	0.060	0.070	0.045	0.054
	R (median)		0.065	0.072	0.038	0.076	0.067	0.047	0.078	0.056	0.056	0.064	0.053	0.054	0.069	0.058	0.047
Wrapped Cauchy	LS		0.019	0.021	0.033	0.025	0.033	0.053	0.021	0.039	0.060	0.024	0.044	0.056	0.034	0.050	0.053
	B (mean)		0.037	0.042	0.050	0.043	0.048	0.064	0.038	0.051	0.065	0.043	0.052	0.056	0.043	0.052	0.050
	B (median)		0.024	0.034	0.017	0.035	0.047	0.050	0.032	0.047	0.058	0.039	0.047	0.056	0.047	0.051	0.050
	R (mean)		0.086	0.091	0.130	0.087	0.077	0.084	0.084	0.063	0.066	0.083	0.056	0.054	0.077	0.055	0.050
	R (median)		0.065	0.067	0.050	0.079	0.067	0.067	0.073	0.061	0.051	0.073	0.049	0.053	0.069	0.053	0.051
Wrapped mixture	LS		0.025	0.031	0.023	0.029	0.053	0.031	0.039	0.070	0.046	0.044	0.066	0.045	0.070	0.064	0.050
	B (mean)		0.046	0.062	0.039	0.046	0.071	0.042	0.066	0.076	0.060	0.065	0.064	0.051	0.086	0.058	0.054
	B (median)		0.031	0.040	0.014	0.046	0.072	0.040	0.048	0.070	0.044	0.063	0.067	0.057	0.060	0.057	0.053
	R (mean)		0.099	0.111	0.105	0.100	0.085	0.063	0.113	0.085	0.060	0.112	0.067	0.047	0.108	0.062	0.050
	R (median)		0.069	0.078	0.044	0.091	0.079	0.054	0.095	0.075	0.045	0.099	0.067	0.055	0.080	0.056	0.053

Table 4.2 Empirical power of the large-sample (LS), bootstrap (B) and randomization (R) versions of the proposed test as a function of the underlying symmetric distribution, ρ and n. The two options used in the symmetrization of the samples are denoted by (mean) and (median). The nominal significance level is $\alpha = 0.05$.

Distribution	Test version	n	10			20			30			50			100		
		ρ	0.20	0.45	0.70	0.20	0.45	0.70	0.20	0.45	0.70	0.20	0.45	0.70	0.20	0.45	0.70
WSNC	LS		0.020	0.024	0.046	0.020	0.015	0.051	0.022	0.051	0.089	0.026	0.064	0.164	0.026	0.090	0.309
$\lambda = 2$	B (mean)		0.044	0.046	0.036	0.038	0.032	0.068	0.044	0.063	0.105	0.040	0.077	0.180	0.037	0.098	0.316
	B (median)		0.026	0.028	0.013	0.031	0.046	0.072	0.035	0.057	0.104	0.037	0.066	0.169	0.036	0.106	0.308
	R (mean)		0.094	0.088	0.088	0.091	0.087	0.087	0.085	0.082	0.116	0.075	0.085	0.182	0.064	0.104	0.323
	R (median)		0.068	0.059	0.038	0.072	0.067	0.083	0.076	0.069	0.108	0.070	0.077	0.170	0.062	0.112	0.310
WSNC	LS		0.016	0.022	0.041	0.026	0.072	0.152	0.020	0.108	0.297	0.032	0.210	0.550	0.049	0.437	0.877
$\lambda = 5$	B (mean)		0.031	0.045	0.074	0.043	0.093	0.194	0.042	0.138	0.337	0.046	0.229	0.575	0.070	0.449	0.886
	B (median)		0.031	0.037	0.027	0.031	0.086	0.181	0.035	0.128	0.327	0.043	0.217	0.563	0.062	0.460	0.892
	R (mean)		0.078	0.093	0.141	0.093	0.122	0.235	0.086	0.163	0.352	0.094	0.244	0.586	0.103	0.457	0.886
	R (median)		0.080	0.073	0.067	0.074	0.110	0.204	0.072	0.150	0.339	0.080	0.247	0.573	0.090	0.472	0.893
WSNC	LS		0.016	0.029	0.045	0.028	0.100	0.227	0.036	0.194	0.424	0.061	0.363	0.736	0.136	0.696	0.972
$\lambda = \infty$	B (mean)		0.041	0.058	0.085	0.048	0.132	0.285	0.060	0.217	0.470	0.088	0.377	0.745	0.161	0.699	0.974
	B (median)		0.032	0.041	0.031	0.040	0.126	0.264	0.049	0.202	0.478	0.071	0.393	0.748	0.141	0.724	0.971
	R (mean)		0.090	0.108	0.163	0.106	0.168	0.345	0.117	0.253	0.491	0.136	0.398	0.755	0.195	0.706	0.973
	R (median)		0.070	0.083	0.074	0.082	0.157	0.296	0.094	0.228	0.500	0.117	0.421	0.759	0.172	0.732	0.972
Wrapped	LS		0.022	0.033	0.070	0.030	0.117	0.309	0.036	0.208	0.539	0.054	0.399	0.783	0.141	0.761	0.980
exponential	B (mean)		0.043	0.064	0.119	0.058	0.151	0.379	0.059	0.237	0.582	0.078	0.415	0.795	0.168	0.769	0.980
	B (median)		0.027	0.041	0.039	0.035	0.141	0.349	0.045	0.232	0.565	0.077	0.410	0.801	0.143	0.776	0.980
	R (mean)		0.098	0.119	0.245	0.113	0.184	0.454	0.111	0.266	0.600	0.122	0.430	0.791	0.208	0.772	0.980
	R (median)		0.067	0.082	0.111	0.080	0.167	0.420	0.088	0.249	0.584	0.114	0.425	0.802	0.172	0.777	0.982

Table 4.1 it can be seen that the randomization versions of the test tend to be non-conservative, particularly for small ρ. We also observe that the ability of the test to maintain the nominal significance level is improved by bootstrapping, particularly for small ρ. Overall, the bootstrap version of the test with $\bar{\theta}$ used to symmetrize samples is identified as maintaining the nominal significance level best. However, there would appear to be little advantage gained from using computer intensive methods for samples in excess of 30 if ρ is moderately large.

From Table 4.2 we observe that the randomization versions of the test are consistently the most powerful. Nevertheless, their use is not to be recommended due to the fact that, as has been established previously, these versions of the test are generally non-conservative. Of the remaining three variants, the bootstrap version with $\bar{\theta}$ used to symmetrize samples is once more identified as having best overall performance. If the results for this bootstrap version of the test are compared with those for its large-sample counterpart, it can be seen that the largest gain in power resulting from bootstrapping is 7%, corresponding to a sample size of $n = 20$. As is to be expected, the gain in power resulting from using the bootstrap is generally greatest for small sized samples.

4.7 Examples

In this section we apply the large-sample and bootstrap versions of the test in the analysis of seven data sets taken from the circular statistics literature.

4.7.1 Azimuths of Palaeocurrents in the Belford Anticline

Fisher (1993, pp. 80-81) tests for the reflective symmetry of 30 cross-bed azimuths of palaeocurrents measured in the Belford Anticline in New South Wales, taken from Fisher & Powell (1989). The data are reproduced in Table 4.3 and represented in the form of a raw circular plot in Figure 4.5.

Table 4.3 Ordered cross-bed azimuths, measured in a clockwise direction from north (in degrees), of 30 palaeocurrents in the Belford Anticline, New South Wales, Australia.

166	177	177	186	214	215	224	224	229	229	232	239	241	242	245	245
250	254	257	257	267	272	277	281	287	290	294	301	315	329		

The analysis of these data given by Fisher is flawed on two counts. Firstly, it makes use of the Wilcoxon signed-rank test after the median direction has been estimated, and secondly, the value of the test statistic, its associated significance probability and the conclusion are awry. The correct p-value for the Wilcoxon test is in fact 0.704, corresponding to a value of the test statistic of 0.38.

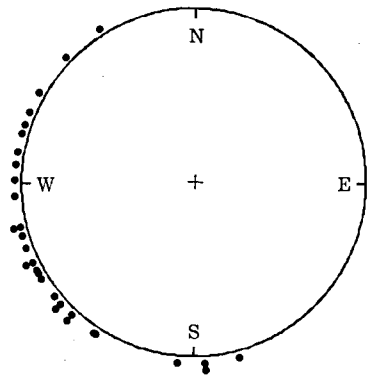

Figure 4.5 Raw circular plot of the cross-bed azimuths of Table 4.3.

For these data the values of $\overline{R}$ and the bias-corrected estimate of ρ calculated using (3.4.1) both equal 0.78 to two decimal places. Using (3.5.4) and (3.5.5), 95% confidence intervals for ρ are, to the same accuracy, (0.68, 0.87) and (0.69, 0.88), respectively. Whilst these estimates correspond to the region of concentration for which the test is relatively powerful, given the small size of the sample, we would not expect the absolute power of the test to be high. In fact, the p-value for the large-sample version of the test was found to be 0.603, and that for the bootstrap version, 0.668. Thus, neither version of the test provides evidence to reject the hypothesis of underlying reflective symmetry. In this and all subsequent references to the bootstrap version of the test, the number of replications used was 2000. Also, in keeping with our findings from the simulaton study of Section 4.6, the sample mean direction, $\overline{\theta}$, was used throughout to symmetrize samples.

4.7.2 Vanishing Angles of Mallard Ducks

Table 1.1 of Mardia (1972, p. 3) reproduces grouped data on the vanishing angles of 714 non-migratory British mallards, taken from Matthews (1961). Over a period of a year, the ducks were taken under sunny conditions from Slimbridge in

Gloucestershire to various sites at distances of between 30 km and 250 km away. When released, the bearing corresponding to the direction at which each bird disappeared from view, the so-called "vanishing angle", was recorded. The bearings were measured clockwise from north. The data as given by Mardia are summarized in Table 4.4. A linear histogram of the data is given in Figure 4.6.

Table 4.4 Vanishing angles (in degrees) of 714 non-migratory British mallards.

Direction	Number of ducks	Direction	Number of ducks
[0–20)	40	[180–200)	3
[20–40)	22	[200–220)	11
[40–60)	20	[220–240)	22
[60–80)	9	[240–260)	24
[80–100)	6	[260–280)	58
[100–120)	3	[280–300)	136
[120–140)	3	[300–320)	138
[140–160)	1	[320–340)	143
[160–180)	6	[340–360)	69

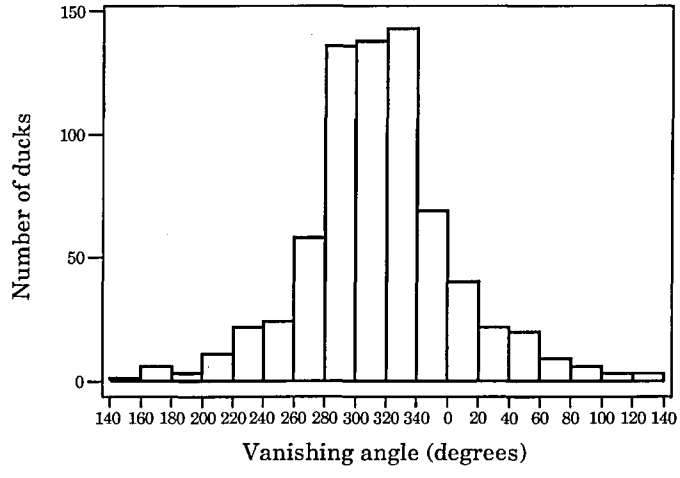

Figure 4.6 Linear histogram of the mallard data of Table 4.4.

Mardia (1972, p. 10) describes the distribution of the data as being "somewhat symmetrical", whilst Mardia & Jupp (1999, p. 5) classify it as being "fairly symmetrical". Without allowing for grouping in any way, and simply centering the observed frequencies in the middle of their respective class intervals, both the value of $\overline{R}$ and the bias-corrected estimate of ρ calculated using (3.4.1) equal 0.72

to two decimal places. The 95% confidence intervals for ρ, calculated using (3.5.4) and (3.5.5), are also the same to two decimal places, corresponding to the interval (0.68, 0.75). Given these estimates of ρ and the large size of the sample, we would not expect the bootstrap version of the test to be any more powerful than the large-sample version. However, as the sample size is so large, we would expect the test to have considerable power. For these data the p-value for the large-sample version of the test is actually 0.124. Thus there exists, in fact, weak to no evidence that the underlying distribution is not reflectively symmetric.

4.7.3 Thunder at Kew

Table 1.3 of Mardia (1972, p. 8) presents grouped data on the frequencies of thunder recorded at Kew during the summers of 1910-1935 for twelve two hour periods of the day. The data were adapted from Bishop (1947). In Table 4.5 we reproduce the data as presented by Mardia. Figure 4.7 provides a graphical representation of the equivalent angular data in the form of a linear histogram.

Table 4.5 Frequencies of occurrence for the 725 occasions on which thunder was recorded at Kew during the summers of 1910-1935.

Time (G.M.T.)	Equivalent Angle (Degrees)	Frequency	Time (G.M.T.)	Equivalent Angle (Degrees)	Frequency
[00–02)	[0–30)	26	[12–14)	[180–210)	133
[02–04)	[30–60)	24	[14–16)	[210–240)	149
[04–06)	[60–90)	14	[16–18)	[240–270)	122
[06–08)	[90–120)	15	[18–20)	[270–300)	80
[08–10)	[120–150)	14	[20–22)	[300–330)	61
[10–12)	[150–180)	65	[22–24)	[330–360)	22

Mardia (1972, p. 10) describes the distribution of these data as being "slightly asymmetrical". Again, without allowing for grouping, and employing the same device of centering the observed frequencies referred to in the previous example, the value of $\overline{R}$ and the bias-corrected estimate of ρ both equal 0.52 to two decimal places. The bias-corrected and non-bias-corrected 95% confidence intervals for ρ are also the same to two decimal places, the common interval being (0.48, 0.57). Whilst these estimates of ρ correspond to an underlying distribution which is quite

dispersed, the sample size of 725 is very large, and hence we would expect the ability of the test to detect asymmetry to be high. We would also expect little benefit to accrue from applying the bootstrap version of the test. In fact, when the large-sample version of the test is applied to these data the p-value obtained is zero to four places of decimal. Thus, our test emphatically rejects underlying reflective symmetry.

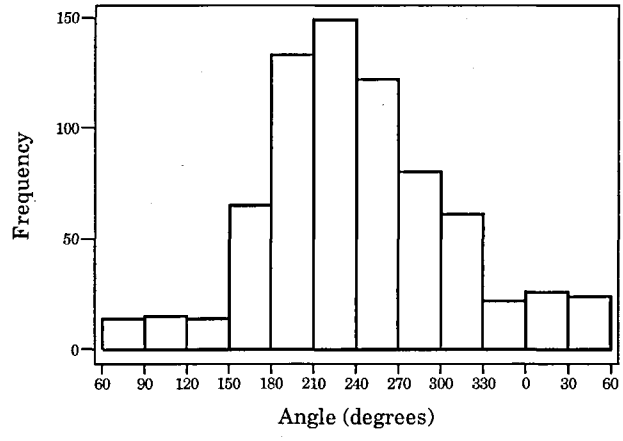

Figure 4.7 Linear histogram of the frequencies of thunder at Kew of Table 4.5.

4.7.4 Excessive Rainfall in the USA

Mardia (1972, p. 8) presents, in his Table 1.4, grouped data on the monthly frequencies of rainfall in excess of 1" or more per hour in the USA for 1908-1937, taken from Dyck & Mattice (1941). The frequencies we consider here are those adjusted for length of month. A summary of these adjusted frequencies is given in Table 4.6, and in Figure 4.8 we represent their distribution in the form of a linear histogram.

Again, Mardia (1972, p. 10) describes these data sets as being "slightly asymmetrical". As the data are grouped, we proceed as in the previous two examples and base our analysis on the centred observed frequencies. For these centred data, the value of $\overline{R}$ and the bias-corrected estimate of ρ both equal 0.55 to two decimal places. The bias-corrected and non-bias-corrected 95% confidence intervals for ρ are also identical to two decimal places, giving the interval (0.54, 0.57). As in the previous example, these estimates correspond to relatively small values of ρ. However, the sample size of 7235 is enormous and consequently we would expect, firstly, the power of the test to be very high and, secondly, the power of the large-sample and bootstrap versions of the test to be similar. In fact, for

these data, the p-value for the large-sample version of the test is, to four decimal places, 0.0004, and so once again we emphatically reject underlying reflective symmetry.

Table 4.6 Adjusted monthly frequencies for the 7235 occasions on which precipitation of 1" or more per hour occurred in the USA between 1908 and 1937.

Month	Equivalent Angle (Degrees)	Frequency (Adjusted)
January	[0–30)	101
February	[30–60)	104
March	[60–90)	231
April	[90–120)	406
May	[120–150)	683
June	[150–180)	1225
July	[180–210)	1475
August	[210–240)	1381
September	[240–270)	907
October	[270–300)	382
November	[300–330)	195
December	[330–360)	145

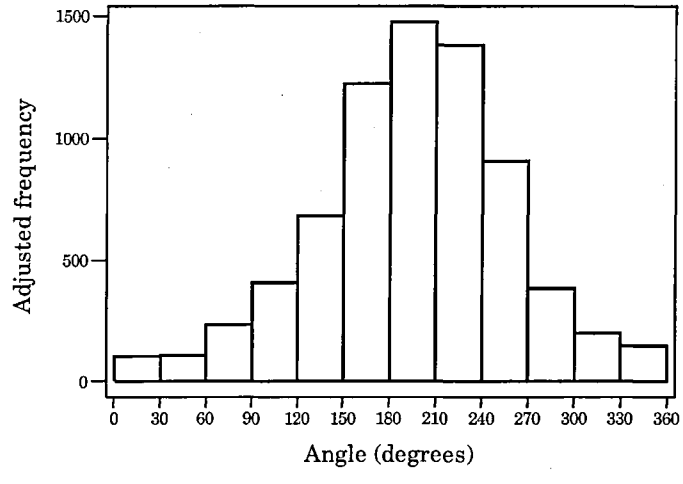

Figure 4.8 Linear histogram of the rainfall data of Table 4.6.

4.7.5 Upper Kamthi River Cross-bed Azimuths

As a last example of a large data set, we consider a sample of grouped data introduced to the literature by Sengupta & Rao (1966). The data consist of the azimuths of 580 cross-beds in the upper Kamthi river in India. The grouped

frequencies for this set of data are reproduced in Table 4.7 and represented graphically in the form of a linear histogram in Figure 4.9.

Table 4.7 Azimuths (in degrees) of 580 cross-beds in the upper Kamthi river, India.

Azimuth (degrees)	Frequency	Azimuth (degrees)	Frequency
[0–20)	75	[180–200)	0
[20–40)	75	[200–220)	21
[40–60)	15	[220–240)	8
[60–80)	25	[240–260)	24
[80–100)	7	[260–280)	16
[100–120)	3	[280–300)	36
[120–140)	3	[300–320)	75
[140–160)	0	[320–340)	90
[160–180)	0	[340–360)	107

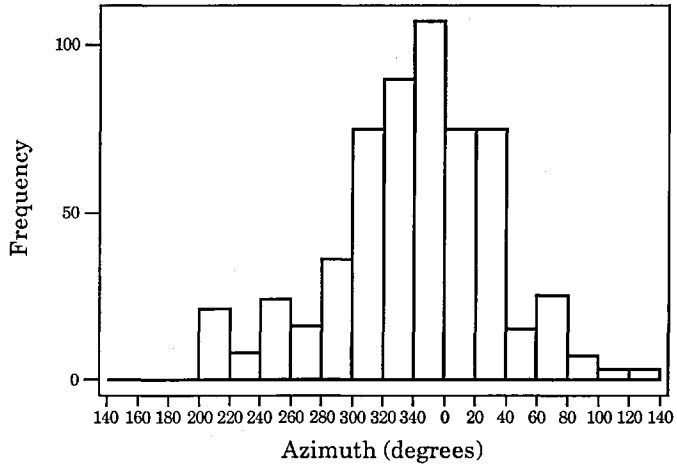

Figure 4.9 Linear histogram of the cross-bed azimuths of Table 4.7.

Mardia (1972, p. 10) describes the distribution of these data as being "asymmetrical". Once more, as the data are grouped, we base our analysis on the observed frequencies centred in the middle of their respective class intervals. For the resulting sample, the value of $\overline{R}$ and the bias-corrected estimate of ρ both equal 0.66 to two decimal places. The 95% confidence intervals for ρ, obtained using (3.5.4) and (3.5.5), are also identical to two decimal places, the common interval being (0.62, 0.70). Again, these estimates correspond to relatively small values of ρ, but as the sample size is 580 we would expect the test to be powerful.

For such a large sample size one would also expect the power of the large-sample and bootstrap versions of the test to be similar. It transpires that for these data the p-value for the large-sample version of the test is, to four decimal places, 0.0014, and so once more we have overwhelming evidence that the parent population is asymmetric. This result partly explains why Sengupta & Rao (1966) found a significant lack of fit when they modelled these data using a von Mises distribution.

4.7.6 Directions of Creek Flow

Fisher (1993, p. 252) presents data on the direction of creek flow at the nearest point to the nests of 50 noisy scrub birds. We reproduce the data values in Table 4.8 and represent their distribution using a raw circular plot in Figure 4.10. The plot gives the impression of a bimodal underlying distribution which is potentially reflectively, but not cyclically, symmetric.

Table 4.8 Ordered directions of creek flow (in degrees) at the nearest point to the nests of 50 noisy scrub birds.

30	90	105	105	105	110	110	125	125	125	125	125	130	130	130	135	135	140
140	140	140	140	145	160	160	160	160	200	200	200	205	215	215	215	225	230
230	240	240	240	240	240	240	250	250	250	250	250	295	295				

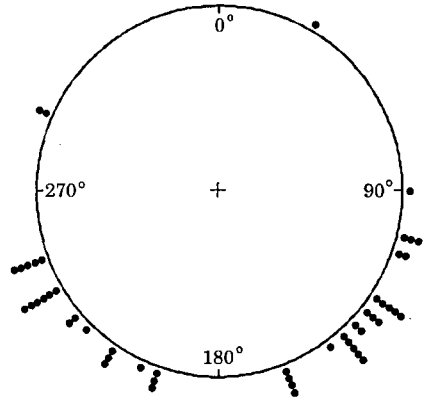

Figure 4.10 Raw circular plot of the creek flow data of Table 4.8.

The mean resultant length and bias-corrected estimate of ρ for these data are 0.55 and 0.54, respectively. In addition, the bias-corrected and non-bias-corrected 95% confidence intervals for ρ are (0.43, 0.64) and (0.44, 0.65), respectively. Whilst

in our Monte Carlo studies we did not investigate the power of the different versions of the test for bimodal distributions, given the results for the unimodal distributions and for sample sizes and ρ-values closest to those for this data set, we would not expect either the large-sample or bootstrap versions of the test to be particularly powerful. In fact, for this sample, the p-values for these two variants of the test were found to be 0.483 and 0.471, respectively. Thus, both versions of the test tend to confirm our earlier impression of an underlying distribution which is reflectively symmetric.

4.7.7 Orientations of Turtles After Egg Laying

In this, our last, example we consider Gould's oft studied turtle data, originally cited in the statistical literature by Stephens (1969). The angles making up the sample correspond to the directions taken by 76 turtles after egg laying. These orientations are given in Table 4.9 and represented in the form of a raw circular plot in Figure 4.11.

Table 4.9 Ordered orientations, measured in a clockwise direction from north (in degrees), of 76 turtles after egg laying.

8	9	13	13	14	18	22	27	30	34	38	38	40	44	45	47	48	48
48	48	50	53	56	57	58	58	61	63	64	64	64	65	65	68	70	73
78	78	78	83	83	88	88	88	90	92	92	93	95	96	98	100	103	106
113	118	138	153	153	155	204	215	223	226	237	238	243	244	250	251	257	268
285	319	343	350														

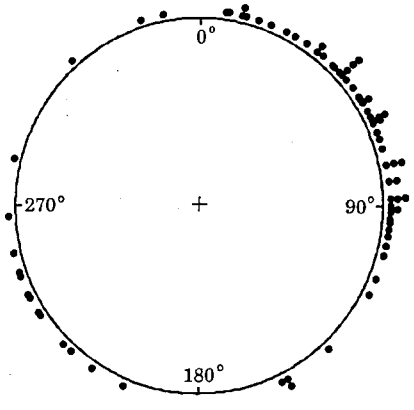

Figure 4.11 Raw circular plot of the turtle data of Table 4.9.

From the circular plot we see that the distribution of the data is bimodal with what appear to be two diametrically opposed modes. Hence, whilst the majority of the turtles move in directions about the primary modal direction, a minority move in directions roughly opposite to those making up the primary mode. It would appear that the underlying distribution could well be reflectively symmetric, but is certainly not 2-fold, that is antipodally, symmetric. The value of $\overline{R}$ and the bias-corrected estimate of ρ for these data are 0.50 and 0.49, respectively. The 95% confidence intervals for ρ, obtained using (3.5.4) and (3.5.5), are identical to two decimal places, the common interval being (0.34, 0.65). Given these estimates of ρ and the moderate size of the sample, we would not expect the test to be particularly powerful. Nor would we expect there to be any great difference between the powers of the large-sample and bootstrap variants of the test. Applying these two versions of the test, we found their corresponding p-values to be 0.744 and 0.758, respectively. Hence, according to both versions of the test, these data provide no evidence against underlying reflective symmetry. If, unnecessarily in this case, the device of m-fold wrapping of the circle is applied, i.e. under the assumption of antipodal symmetry we double each data value, then the p-values for the same two versions of the test are found to be 0.655 and 0.671, respectively.

4.8 Summary and Directions for Future Research

In this last section of the chapter we summarize the main issues which have been addressed within it so far, and outline some potential avenues for future related research.

4.8.1 Summary

In this chapter we have addressed the testing of circular data for reflective symmetry about an unknown central direction, against asymmetry. We reviewed the background to the problem in Section 4.2 and proposed a new large-sample omnibus test in Section 4.3.

The operating characteristics of the test were explored in Section 4.4. In Section 4.4.1 we gave the theoretical asymptotic power of the test, whilst in Section 4.4.2 we presented results from a Monte Carlo experiment designed to explore the test's ability to maintain the nominal significance level and its power. In the latter section we also compared the theoretical and empirical power of the test for four

parent populations of varying skewness. The new test was found, in general, to be conservative, its ability to maintain the nominal significance level improving with increasing sample size and concentration of the underlying distribution. For a finite fixed sample size, the power of the test as a function of ρ tends to improve with increasing concentration before reaching a maximum and subsequently diminishing as ρ approaches its limiting value of 1. For $\rho = 1$, of course, $\bar{b}_2$ is identically zero.

When describing the background to the Monte Carlo study of Section 4.4.2 we also presented results for the fundamental properties of three models which are new to the literature: the wrapped Laplace distribution, a wrapped normal and uniform mixture distribution and the wrapped exponential distribution.

Randomization and bootstrap versions of the test were introduced in Section 4.5. In Section 4.6 we presented results obtained from a further simulation study conducted in order to compare the operating characteristics of these new versions of the test with those of the original large-sample version. On the basis of the admittedly limited results from this study, the bootstrap variant of the test, with $\bar{\theta}$ used in the symmetrization of the data, was identified as the version of the test with best overall performance. However, for samples of size 30 or more, drawn from parent populations which are not highly dispersed, the improvement in performance of the bootstrap version of the test over its large-sample counterpart is only very marginal.

Finally, in Section 4.7, we illustrated the use of the large-sample and bootstrap versions of the new omnibus test, presenting analyses of seven data sets taken from the circular statistics literature.

4.8.2 Directions for Future Research

The test procedures developed in this chapter follow naturally from the use of $\bar{b}_2$ as a fundamental measure of circular skewness. Our derivation of the large-sample version of the omnibus test therefore parallels closely that used by Gupta (1967) in deriving an analogous test for linear data based on the coefficient of skewness $g_1 = m_3 / m_2^{3/2}$. However, $\bar{b}_2$ is only one amongst a range of measures which might be used as the basis of a test for symmetry about an unknown central direction. In this regard, a set of tests which suggests itself for future investigation is that formed by the circular analogues of procedures which have been found to be

powerful tests of symmetry for linear data. The findings of Cabilio & Masaro (1996) are highly relevant in terms of the identification of the members of this set. These two authors used Monte Carlo methods to compare the small-sample operating characteristics of numerous tests of symmetry about an unknown centre. Of the 3tests considered by them, the three tests which performed best were: a test proposed by them based on the difference between the sample mean and median; the test of Boos (1982) based on a minimum Cramér-von Mises distance characterization of a variant of the Hodges-Lehmann location estimator; the test of Finch (1977) based on the relative magnitude of the gaps between successive ordered observations. However, the development of circular equivalents of the last two of these tests would appear to be no easy task. The test of Boos (1982) uses a location estimator based on the median of Walsh averages and this immediately raises problems in terms of deriving an analogous circular test. The test of Finch (1977) requires an ordering of the observations, for which, at least for circular data distributed over the complete range of the unit circle, there is no unique circular equivalent. The development of a test for circular data analogous to that of Cabilio & Masaro (1996), based on the statistic $\bar{\theta} - \tilde{\theta}$, perhaps suggests itself as being potentially more tractable. Nevertheless, the use of $\tilde{\theta}$ raises certain practical difficulties because, as Fisher (1993, p.35) points out, for real data the median direction need not be uniquely defined, particularly if the data are multimodal or isotropic. In addition, to our knowledge, nothing appears to have been published regarding the joint distribution of $\bar{\theta}$ and $\tilde{\theta}$. Should the distribution theory of a test based on some pivotal statistic involving $\bar{\theta} - \tilde{\theta}$ prove too daunting, one could always define a computer intensive version of it in which the double bootstrap was used to approximate the variance of $\bar{\theta} - \tilde{\theta}$.

As we explained in Section 4.2, when the data in a sample are distributed over a reduced arc of the circle, one can analyze them as if they were linear after applying a suitable rotation and unwrapping them to the line. The three tests of symmetry referred to in the previous paragraph are powerful tests whose use should be contemplated in such a context. We note, however, that the recommendations regarding the application of the three tests are based on their performance for data from unimodal populations only.

Related to this last issue, our empirical investigations of the operating characteristics of the various versions of our test based on $\bar{b}_2$ were also rather limited. It would be of interest to explore their performance for other unimodal distributions, as well as for multimodal distributions, and for a more complete range of ρ-values. Given the experience gained from our small scale Monte Carlo experiment, the computing resources required to conduct such an in-depth investigation would be considerable. It would also be interesting to compare the operating characteristics of our test with those of linear tests of symmetry, for situations in which both can be applied. From such an investigation one could potentially draw up recommendations for the use of the two forms of test based on the mean resultant length of a sample. We note that an investigation along similar lines is presented in the following chapter.

Testing linear data for symmetry about an unknown centre is well known to be a thorny problem, the null and alternative hypotheses embracing all possible forms of underlying distribution. It is therefore not surprising that many of the tests of symmetry proposed in the literature have low power. The compactness of the circle does nothing to improve matters, as most importantly it leads to the loss of the notion of "scale". Consequently, as we have seen, any assessment of the performance of a test of reflective symmetry for circular data must take account of the concentration of the parent population. One potential means of wresting power from where perhaps it should not be wrested is to employ an approach similar to that advocated by Kappenman (1988), i.e. convert the problem of testing the non-parametric hypothesis of symmetry into a parametric one, with the likelihood ratio criterion being used to choose between a member of a rich symmetric class of distributions and a member of a rich asymmetric class. Whilst considering that such an approach has its merits, we have shied away from pursuing it as we view the definition of sufficiently rich classes of circular distributions, capable of adequately modelling the wide range of possible empirical distributions of circular data, as being an insurmountable task. Should one be prepared to restrict one's attention to unimodal distributions only, such a task might appear less ambitious.

Chapter 5 Testing for Circular Reflective Symmetry About a Known or Specified Median Axis

5.1 Introduction

In Chapter 4 we considered procedures for testing circular data for the null hypothesis of an underlying distribution which is reflectively symmetric about an unspecified central direction against the alternative hypothesis of a parent population which is asymmetric. For the most part, the present chapter is devoted to the consideration of procedures designed for testing the null hypothesis of an underlying circular distribution which is reflectively symmetric about a known median axis against the alternative hypothesis that the underlying distribution is skew. To our knowledge, nothing has previously been published regarding the testing of this pair of hypotheses.

In Section 5.2 we review the background to testing symmetry about a known or specified centre, in the case of data on the line, and median axis for data on the circle. In Section 5.3 we introduce two new omnibus procedures for testing for circular reflective symmetry about a known median axis against skew alternatives. In addition, in Section 5.4 we describe the circular analogues of three linear tests which can also be employed in this testing set-up.

Theoretical results for the asymptotic power of the two new test procedures are derived in Section 5.5, and in Section 5.6 we present the details of a Monte Carlo experiment designed to explore and compare the operating characteristics of the two new tests and those of the circular analogues of the three linear tests. The empirical results from this simulation study are used to establish, in Section 5.6.3, a simple strategy for testing circular data from unimodal populations for reflective symmetry about a known median axis against skew alternatives.

In Section 5.7 we change tack and consider the use of the two new procedures as tests of symmetry about a specified median axis against rotation alternatives. We derive results for the asymptotic power of the two tests against rotation alternatives in Section 5.7.1. In Section 5.7.2 we present the details of a simulation

study designed to compare the small-sample power characteristics of the two new tests with those of the circular analogues of four linear tests.

In Section 5.8 we discuss the use and limitations of the various test procedures considered in the previous sections. So as to illustrate the application of the testing strategy developed in Section 5.6.3, in Section 5.9 we present analyses of four data sets collected during animal orientation experiments.

The chapter closes, in Section 5.10, with a summary of its content and a brief indication of potential lines of related future research.

5.2 Background to the Two Types of Testing Problem for Data on the Line and Circle

In the statistical literature there exists a vast array of procedures designed for testing linear data for symmetry when the median of the underlying distribution is assumed to be "known" or when the centre of the distribution is "specified". Some of these procedures can be applied in both testing contexts, while others are designed specifically for testing one or other of them.

For the set-up in which the median of the underlying distribution is assumed to be known, the relevant procedures are used to explore the hypothesis that the parent population is symmetric about the known median against the alternative that the underlying distribution is skew. Henceforth we refer to these procedures as tests of symmetry about a known median against skew alternatives.

In the testing scenario in which the centre of the distribution is specified, the underlying distribution is assumed to be symmetric and interest focuses on whether the centre of the distribution is correctly specified or not. These latter tests are generally referred to as tests of symmetry against location shift alternatives. The best known tests of this type are the sign test, the one-sample Wilcoxon test and the normal-scores test of Fraser (1957). The sign test is known to be the locally most powerful linear rank test against location shift alternatives, as well as being asymptotically optimum, when the underlying distribution is Laplace. Similarly, the Wilcoxon and normal-scores tests are the locally most powerful linear rank tests, as well as being asymptotically optimum, when the parent population is logistic or normal, respectively. See, for example, Hájek & Sidák (1967, Section III.5.1).

However, there are numerous other procedures which can be used as omnibus tests of symmetry against location shift alternatives. Amongst these figure the test

of Smirnov (1947) and Butler (1969) based on the Kolmogorov-Smirnov statistic, and the tests of Orlov (1972), Rothman & Woodroofe (1972), Srinivasan & Godio (1974) and Hill & Rao (1977) which employ various Cramér-von Mises statistics. These empirical distribution function based procedures can also be used as tests of symmetry about a known median against skew alternatives. So can the Wilcoxon test, but clearly the sign test cannot usefully be applied in this testing scenario. In recent times, a class of procedures has been developed exclusively for testing for symmetry about a known median against skew alternatives. Cohen & Menjoge (1988) and McWilliams (1990) independently developed a test based on the number of runs about the known median. A modification of this test, which incorporated the percentile modified test scores of Gastwirth (1965), was considered by Modarres & Gastwirth (1996). Tajuddin (1994) proposed a conditional test based on the Wilcoxon two-sample test, and Modarres & Gastwirth (1998) introduced a hybrid test involving a combination of the sign test and a percentile modified two-sample Wilcoxon test.

The development of equivalent procedures for testing *circular* data for reflective symmetry about a known or specified median axis has received considerably less attention in the literature. Schach (1969) considered the equivalent of tests of symmetry about a specified median and obtained results for locally most powerful rank procedures for testing circular data for reflective symmetry about a specified median axis against rotation alternatives. In practice, suitably adapted versions of the sign and one-sample Wilcoxon tests are usually employed as omnibus procedures in this testing context; see, for example, Mardia & Jupp (1999, Section 8.2.1). These versions of the Wilcoxon and sign tests are described in Sections 5.4.1 and 5.7, respectively.

As far as we are aware, nothing has previously been published regarding procedures for testing circular data for reflective symmetry about a known median axis against the alternative of an underlying distribution which is skew. This is somewhat surprising, as in the modelling of circular data interest will often focus on whether the underlying distribution from which the data arise is symmetric about an axis associated with the experimental set-up under consideration. For instance, in many animal orientation experiments the variable of interest is a circular one, namely the direction in which an animal moves in response to a stimulus, such as a source of light, heat or sound. A basic question of scientific interest in such contexts is whether the animals move in directions described by a circular distribution which is symmetric about a particular axis associated with the

orientation of the stimulus, or whether the distribution about the same axis is skew.

Notwithstanding the above, if the data under consideration are concentrated on a reduced arc of the unit circle then they can effectively be treated as being linear. For the analysis of data of this type, the results from simulation experiments reported by McWilliams (1990) and Modarres & Gastwirth (1996, 1998) are of particular relevance. These results show that, for linear data from a wide range of unimodal asymmetric distributions from the generalized lambda family of Ramberg & Schmeiser (1974), the modified runs and hybrid tests of Modarres & Gastwirth (1996, 1998) can be the most powerful, nominal 5% sized, procedures known for testing for symmetry against skew alternatives. Thus, for data distributed over a reduced arc of the unit circle, one would expect the circular analogues of these procedures to be powerful tests of circular reflective symmetry about a known median axis against skew alternatives. However, for angular data which are distributed over a considerable arc of the unit circle, which we will call truly circular data, one might envisage the power of such tests to be considerably lower.

In the following section we introduce two simple omnibus procedures designed to test truly circular data for reflective symmetry about a known median axis against skew alternatives. In Section 5.4 we describe the circular analogues of the one-sample Wilcoxon, runs and modified runs tests which might be employed in the same testing context. The simulation results reported in Modarres & Gastwirth (1998) indicate that, for appropriate combinations of its parameters, α_1 , α_2 and p, their hybrid test can sometimes be even more powerful than their modified runs test. Nevertheless, we do not consider its circular analogue due to what we view as a major shortcoming in the original specification of the test. Although Modarres & Gastwirth (1998) offer advice on the choice of p, they only present simulation results for the combination $\alpha = 0.05$, $\alpha_1 = 0.01$ and $\alpha_2 = 0.0404$, and offer only very imprecise advice as to how α_1 and α_2 should be chosen for general values of α. Given this lack of specification, we fear that the use of their hybrid procedure has great potential for converting the original problem into one irrevocably confounded by multiple testing.

5.3 Two Asymptotically Distribution-free Procedures for Testing for Circular Reflective Symmetry About a Known Median Axis

In this section we introduce two new omnibus procedures for testing for circular reflective symmetry about a known median axis against skew alternatives. The first is based on a statistic similar to $\bar{b}_2$ which we used in Chapter 4 as the basis of a test of reflective symmetry about an unknown central direction against asymmetry. The second uses the difference between the sample mean direction and the known median direction as a measure of skewness. As we show, both tests are asymptotically distribution-free.

5.3.1 An Omnibus Test Based Upon the Second Sine Moment About a Known Median Direction

In what follows we suppose that $\tilde{\mu}$ is a known median direction, so that a median axis is known to pass through $\tilde{\mu}$ and $\tilde{\mu}+\pi$. The first of our proposed tests is based upon the statistic $b_2^* = \frac{1}{n}\sum_{i=1}^{n} \sin 2(\theta_i - \tilde{\mu})$, the sample second sine moment about the median direction $\tilde{\mu}$. This statistic is closely related to the sample second sine moment about the mean direction, $\bar{b}_2 = \frac{1}{n}\sum_{i=1}^{n} \sin 2(\theta_i - \bar{\theta})$, used as the basis of the test procedure considered in Chapter 4. The following theorem gives the asymptotic distribution of b_2^* under quite general conditions.

Theorem 5.1 *Let $\theta_1,...,\theta_n$ be n independently and identically distributed random variables from an angular distribution for which $\rho \neq 1$ and $\tilde{\mu}$ is known to be a median direction. Then, as $n \to \infty$,*

$$\frac{b_2^* - E(b_2^*)}{\{\mathrm{var}(b_2^*)\}^{1/2}} \xrightarrow{D} \mathrm{N}(0,1), \tag{5.3.1}$$

where $E(b_2^*) = \beta_2^*$,

$$\mathrm{var}(b_2^*) = \frac{1}{2n}\{1 - \alpha_4^* - 2(\beta_2^*)^2\}, \tag{5.3.2}$$

and $\alpha_p^* = E\{\cos p(\theta - \tilde{\mu})\}$ and $\beta_p^* = E\{\sin p(\theta - \tilde{\mu})\}$ are the pth cosine and sine moments of the random angle θ about $\tilde{\mu}$, respectively.

Proof Given the definition of b_2^*, and denoting $(\theta_i - \tilde{\mu})$ by θ_i^*,

$$E(b_2^*) = E\{\sin 2(\theta_i - \tilde{\mu})\} = E(\sin 2\theta_i^*) = \beta_2^*$$

and

$$\begin{aligned}
\text{var}(b_2^*) &= \frac{1}{n}\text{var}(\sin 2\theta_i^*) \\
&= \frac{1}{n}\left[E(\sin^2 2\theta_i^*) - \{E(\sin 2\theta_i^*)\}^2\right] \\
&= \frac{1}{n}\left[\frac{1}{2}\{1 - E(\cos 4\theta_i^*)\} - (\beta_2^*)^2\right] \\
&= \frac{1}{n}\left\{\frac{1}{2}(1-\alpha_4^*) - (\beta_2^*)^2\right\} \\
&= \frac{1}{2n}\{1 - \alpha_4^* - 2(\beta_2^*)^2\}
\end{aligned}$$

The asymptotic normality of b_2^* follows from the central limit theorem, and hence we obtain (5.3.1) under the assumption that the underlying distribution of θ is not a point distribution, i.e. $\rho \neq 1$.

For a distribution which is symmetric about $\tilde{\mu}$, $f(\theta) = f(2\tilde{\mu} - \theta)$ for all $\theta \in [0, 2\pi)$ and $\beta_2^* = 0$. Of course, $f(\theta)$ is also symmetric about the antipodal direction $\tilde{\mu} + \pi$. An estimate of the variance of b_2^* under these conditions can be obtained using (5.3.2) with $\beta_2^* = 0$ and α_4^* replaced by the consistent plug-in estimate $a_4^* = \frac{1}{n}\sum_{i=1}^{n}\cos 4(\theta_i - \tilde{\mu})$.

Given the above, we define the following omnibus test for circular reflective symmetry about $\tilde{\mu}$.

The b2-star Test

Calculate the studentized statistic

$$b_2^* / \{\hat{\text{var}}(b_2^*)\}^{1/2}, \tag{5.3.3}$$

where $\hat{\text{var}}(b_2^*) = \frac{1}{2n}(1 - a_4^*)$. Circular reflective symmetry about $\tilde{\mu}$ is rejected in favour of an underlying distribution which is skew, at the $100\alpha\%$ level, if the

absolute value of (5.3.3) is greater than $z_{\alpha/2}$, the upper $\alpha/2$ quantile of the standard normal distribution. In the sequel we refer to this procedure as the b2-star test.

As has already been noted, for an underlying distribution which is symmetric about $\tilde{\mu}$, $E(b_2^*) = \beta_2^* = 0$. However, here we make the additional observation that for an underlying distribution which is asymmetric the expectation of b_2^* is also 0 when $\beta_2^* = \beta_2 \cos 2\tilde{\mu} - \alpha_2 \sin 2\tilde{\mu} = 0$, i.e. when $\tan 2\tilde{\mu} = \beta_2 / \alpha_2$. In order to illustrate when this occurs for a highly asymmetric unimodal distribution, consider the wrapped exponential distribution. Using the results from Section 4.4.2.2.1 of Chapter 4, $\beta_2^* = 0$ for the wrapped exponential distribution when

$$-\frac{2}{\lambda}\log_e\left\{\frac{1-e^{-2\pi\lambda}}{2\left(1-e^{-\pi\lambda}\right)}\right\} - \tan^{-1}\left(\frac{2}{\lambda}\right) = 0.$$

Solving numerically for λ, one obtains $\lambda = 1.44337$ which corresponds to $\rho = 0.82199$. Hence, if the underlying distribution is wrapped exponential we would expect the power of the b2-star test to be low for ρ-values in the neighbourhood of 0.82199.

5.3.2 An Omnibus Test Based Upon the Difference Between the Sample Mean Direction and the Known Median Direction

For our second test we make an additional assumption regarding the underlying distribution, namely that the known median direction, $\tilde{\mu}$, is the unique median direction of the distribution. If this is the case then $\rho \neq 0$ and the underlying distribution also has a unique mean direction, μ. If, in addition, the parent population is symmetric then $\tilde{\mu} = \mu$, and, estimating μ using $\bar{\theta}$, the difference $\bar{\theta}^* = \bar{\theta} - \tilde{\mu}$, measured as an angle in the interval $[-\pi, \pi)$, suggests itself as a potential measure of skewness. From Theorem 3.1 of Chapter 3 we have

$$E(\bar{\theta}) = \mu - \frac{\bar{\beta}_2}{2n\rho^2} + O(n^{-3/2}),$$

$$\text{var}(\bar{\theta}) = \frac{(1-\bar{\alpha}_2)}{2n\rho^2} + O(n^{-3/2}),$$

and that the asymptotic distribution of $\bar{\theta}$ is normal. For a symmetric parent population, $\bar{\beta}_2 = \beta_2^* = 0$ and $\bar{\alpha}_2 = \alpha_2^*$. Estimating ρ using $\bar{R}$, and $\bar{\alpha}_2$ using $a_2^* = \frac{1}{n}\sum_{i=1}^{n} \cos 2(\theta_i - \tilde{\mu})$, then, for a symmetric underlying distribution, as $n \to \infty$

$$\frac{\bar{\theta}^*}{\{\hat{\text{var}}(\bar{\theta})\}^{1/2}} \xrightarrow{D} \text{N}(0,1), \tag{5.3.4}$$

where $\hat{\text{var}}(\bar{\theta}) = \frac{1}{2n\bar{R}^2}(1-a_2^*)$.

Given the distributional result in (5.3.4), we define the second of our asymptotically distribution-free tests of circular reflective symmetry about $\tilde{\mu}$ against skew alternatives as follows.

The Theta-bar Test

Calculate the studentized statistic

$$\frac{\bar{\theta}^*}{\{\hat{\text{var}}(\bar{\theta})\}^{1/2}}. \tag{5.3.5}$$

Circular reflective symmetry about $\tilde{\mu}$ is rejected in favour of an underlying distribution which is skew, at the $100\alpha\%$ level, if the absolute value of (5.3.5) is greater than $z_{\alpha/2}$. In what follows we refer to this procedure as the theta-bar test.

5.4 Circular Analogues of Three Linear Tests

Circular analogues of the standard and most powerful procedures for testing linear data for symmetry against skew alternatives can also be used to test for circular reflective symmetry about a known median axis against skew alternatives. For a known median direction, $\tilde{\mu}$, one first calculates the differences $\theta_i^* = \theta_i - \tilde{\mu}$ as angles in the interval $[-\pi, \pi)$, a transformation of the data equivalent to that described in Mardia & Jupp (1999, Section 8.2.1). In the descriptions which follow we assume the underlying distribution to be continuous.

5.4.1 The One-sample Wilcoxon Test

For the circular version of the one-sample Wilcoxon test, let $\theta^*_{(1)},...,\theta^*_{(n)}$ denote the θ^*_i values ordered from smallest to largest according to absolute value. Encode the signs of the $\theta^*_{(i)}$ using the indicator variables $S_1,...,S_n$ where

$$S_i = \begin{cases} 1 & \text{if } \theta^*_i \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

After assigning the ranks $1,...,n$ to the $\theta^*_{(i)}$, the test statistic is simply $W = \sum_{i=1}^{n} iS_i$.

Circular symmetry is rejected if W is judged to be extremely large or small when compared with the usual critical values of the test.

5.4.2 The Runs Test

In order to conduct the circular version of the runs test of Cohen & Menjoge (1988) and McWilliams (1990), calculate $R = \sum_{i=2}^{n} I_i$, where

$$I_i = \begin{cases} 0 & \text{if } S_i = S_{i-1} \\ 1 & \text{otherwise} \end{cases}.$$

Circular symmetry is rejected if R is judged to be extremely small in comparison with the critical values of the binomial distribution with parameters $n-1$ and $1/2$.

5.4.3 The Modified Runs Test

For the circular version of the modified runs test of Modarres & Gastwirth (1996), calculate $M_p = \sum_{i=np+2}^{n} \phi(i) I_i$, where

$$\phi(i) = \begin{cases} i - np & \text{if } i > np \\ 0 & \text{otherwise} \end{cases},$$

p is a trimming proportion and, for simplicity of notation, np is assumed to be an integer. The $\phi(i)$ are the percentile modified test scores of Gastwirth (1965). The large-sample version of the test rejects circular symmetry if the value of $\{M_p - \mu(M_p)\}/\sigma(M_p)$ is judged to be extremely large and negative when compared with the quantiles of the standard normal distribution, where

$$\mu(M_p) = \{n(1-p)-1\}\{n(1-p)+2\}/4, \text{ and}$$

$$\sigma^2(M_p)=\{n(1-p)-1\}\{2n^2(1-p)^2+5n(1-p)+6\}/24.$$

The results presented in Modarres & Gastwirth (1998) indicate that the choice p = 0.6 leads to a powerful nominal 5% sized test of symmetry for linear data when the underlying distribution is unimodal and skew.

5.5 The Asymptotic Power of the Two New Tests Against Skew Alternatives

5.5.1 Power of the b2-star Test

Using the results for the distribution of b_2^* given in the proof of Theorem 5.1, under the null hypothesis of an underlying distribution which is symmetric about $\tilde{\mu}$, the asymptotic distribution of b_2^* is normal with mean 0 and variance

$$\text{var}_0(b_2^*)=\frac{1}{2n}(1-\alpha_4^*).$$

Under the alternative hypothesis of an underlying distribution which is skew the asymptotic distribution of b_2^* is normal with mean $E_1(b_2^*)=\beta_2^*$ and variance

$$\text{var}_1(b_2^*)=\frac{1}{2n}\{1-\alpha_4^*-2(\beta_2^*)^2\}.$$

Thus, for a significance level of $100\alpha\%$, the asymptotic power of the b2-star test against skew alternatives is given by

$$1-\Phi\left(\frac{z_{\alpha/2}\{\text{var}_0(b_2^*)\}^{1/2}-\beta_2^*}{\{\text{var}_1(b_2^*)\}^{1/2}}\right)+\Phi\left(\frac{-z_{\alpha/2}\{\text{var}_0(b_2^*)\}^{1/2}-\beta_2^*}{\{\text{var}_1(b_2^*)\}^{1/2}}\right). \qquad (5.5.1)$$

The form taken by (5.5.1) as a function of ρ is exhibited in Figure 5.3 for the same four skew distributions used in Chapter 4.

5.5.2 Power of the Theta-bar Test

From the results for the asymptotic distribution of $\bar{\theta}$ given in Section 5.3.2, under the null hypothesis of an underlying distribution which is symmetric about the median direction $\tilde{\mu}$, the asymptotic distribution of $\bar{\theta}$ is normal with mean $\tilde{\mu}$ and variance

$$\text{var}_0(\bar{\theta})=\frac{(1-\alpha_2^*)}{2n\rho^2}.$$

Under the alternative hypothesis of an underlying distribution which is skew the asymptotic distribution of $\bar{\theta}$ is normal with mean

$$E_1(\bar{\theta}) = \mu - \frac{\bar{\beta}_2}{2n\rho^2}$$

and variance

$$\text{var}_1(\bar{\theta}) = \frac{(1-\bar{\alpha}_2)}{2n\rho^2}.$$

It follows that, for a significance level of $100\alpha\%$, the asymptotic power of the test against skew alternatives is given by

$$1-\Phi\left(\frac{z_{\alpha/2}\{\text{var}_0(\bar{\theta})\}^{1/2} - E_1(\bar{\theta}^*)}{\{\text{var}_1(\bar{\theta})\}^{1/2}}\right) + \Phi\left(\frac{-z_{\alpha/2}\{\text{var}_0(\bar{\theta})\}^{1/2} - E_1(\bar{\theta}^*)}{\{\text{var}_1(\bar{\theta})\}^{1/2}}\right), \quad (5.5.2)$$

where $E_1(\bar{\theta}^*) = E_1(\bar{\theta}) - \tilde{\mu}$ measured as an angle in the interval $[-\pi, \pi)$. The form taken by (5.5.2) as a function of ρ is illustrated in Figure 5.4.

5.6 Monte Carlo Investigation of the Operating Characteristics of the Tests Against Skew Alternatives

5.6.1 Design of the Simulation Experiment

In order to make recommendations regarding the potential uses of our two new tests and the circular analogous of the three linear tests considered in Section 5.4, an assessment is required of their ability to maintain the nominal significance level for symmetric parent populations, and their power against asymmetric alternatives. So as to investigate these two operating characteristics of the various tests for samples drawn from unimodal populations, we conducted a simulation study.

The unimodal distributions used in the study were the same ones as employed in Chapter 4, namely the wrapped normal, wrapped Laplace, wrapped Cauchy, wrapped normal and uniform mixture, wrapped exponential and wrapped skew-normal distributions, the latter with values of the skewness parameter, λ, equal to 2, 5, and ∞. Plots of the densities of these various distributions were given in Figures 4.1 and 4.2 of Chapter 4. In addition, the same sample sizes, ρ-values and nominal significance levels as described in Section 4.4.2.3 were used. However, in the study presently under consideration, the asymmetric distributions were centred so as to have the zero direction as their median direction, and the number of

samples simulated for each distribution, sample size and ρ combination was 5000. In keeping with the findings of Modarres & Gastwirth (1998), we limited our attention to the asymptotic version of the modified runs test with $p = 0.6$. The Wilcoxon and runs tests were randomized so as to have sizes as exactly specified.

5.6.2 Results

The results obtained for the two operating characteristics of the various tests are presented separately in the following two subsections. Once again, only the results for a nominal significance level of 5% are given, the conclusions drawn from them also being representative of those for the other two significance levels considered in the study.

5.6.2.1 Ability to Maintain the Nominal Significance Level

In Figures 5.1 and 5.2 we display the results for the estimated size of the b2-star and theta-bar tests, respectively, for the four symmetric distributions used in the study. As in Chapter 4, the results for sample sizes 30 and 200 have been omitted from these figures so as not to obscure their content. Given that 5000 simulated samples were used to obtain each of the empirical results represented in these plots, the standard error of any such result is approximately 0.003.

It can be seen from Figure 5.1 that the b2-star test's ability to maintain the nominal significance level is very good for all the various distribution, sample size and ρ combinations considered. The deviations about the nominal significance level are what might be expected of the empirical significance level of a truly 5% sized test under random variation.

Turning to Figure 5.2, we see that the theta-bar test's ability to hold the nominal level is not nearly as good. For small ρ-values the test is non-conservative, particularly for small-sized samples. This behaviour is perhaps to be expected given that as $\rho \to 0$ all four distributions tend to the uniform distribution and the variabilty of $\bar{\theta}^*$ increases greatly, particularly for small-sized samples. As one might also have envisaged, the test's non-conservatism for small ρ-values diminishes with increasing sample size.

The circular analogues of the three linear tests all maintained the nominal significance well across the complete range of distribution, sample size and ρ combinations considered.

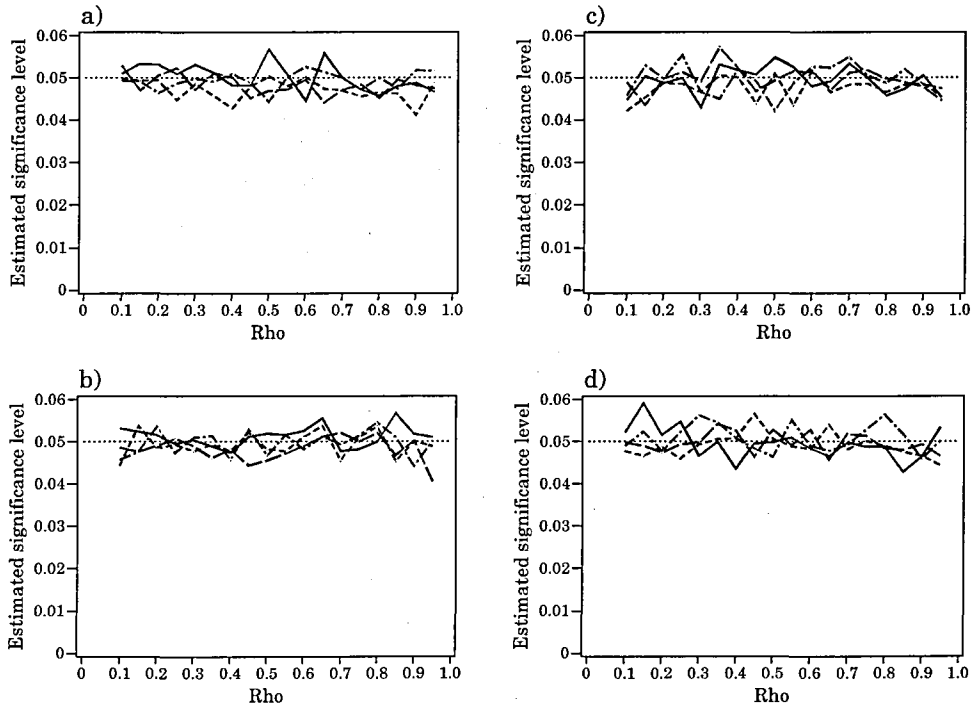

Figure 5.1 Estimated size of the b2-star test for the distributions: a) wrapped normal, b) wrapped Laplace, c) wrapped Cauchy, d) wrapped normal and uniform mixture. Sample sizes represented are: - - - - - (n = 20); — — — (n = 50); — · — (n = 100); ——— (n = 500). The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$.

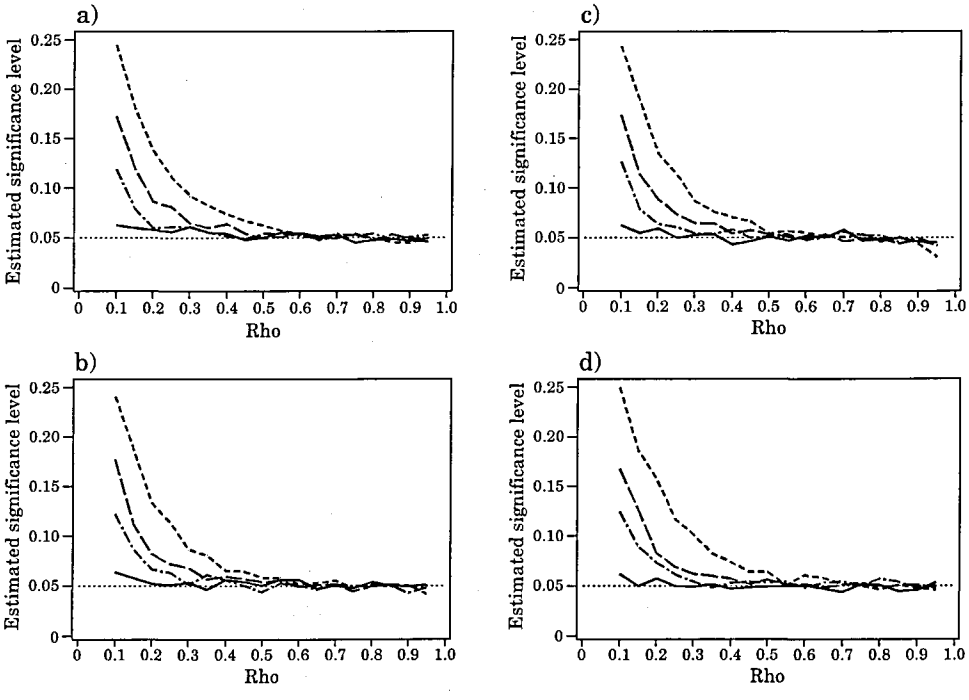

Figure 5.2 Estimated size of the theta-bar test for the distributions: a) wrapped normal, b) wrapped Laplace, c) wrapped Cauchy, d) wrapped normal and uniform mixture. Sample sizes represented are: - - - - - (n = 20); — — — (n = 50); — · — (n = 100); ——— (n = 500). The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$.

5.6.2.2 Power Against Skew Alternatives

Displayed in Figures 5.3 and 5.4 are the empirical power results for the b2-star and theta-bar tests, respectively. The standard error of any empirical result represented in these two figures is, at most, 0.007.

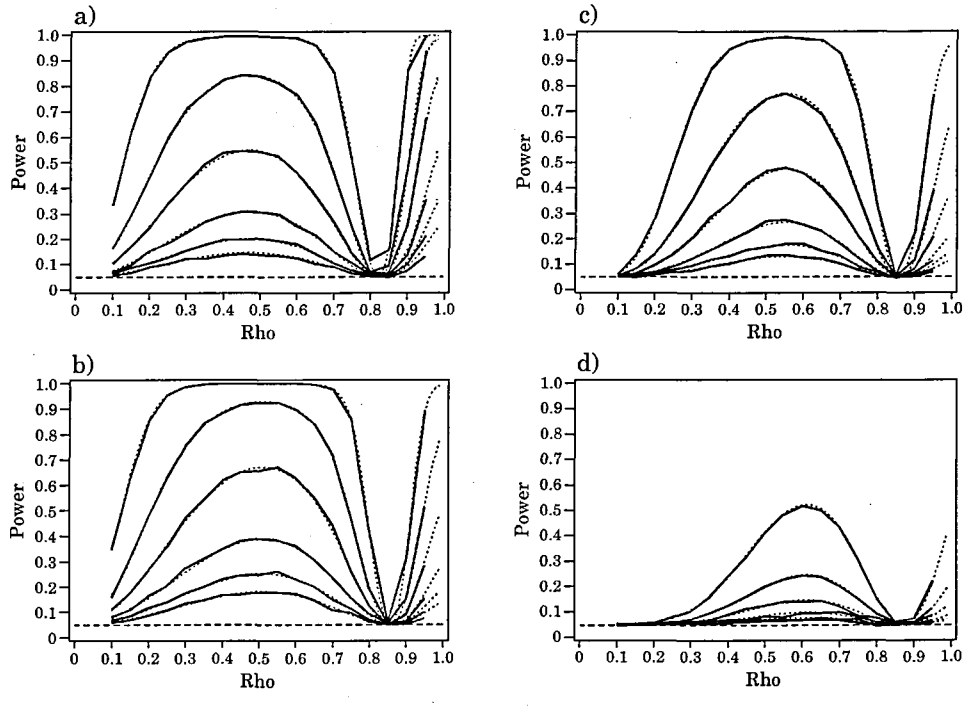

Figure 5.3 Theoretical asymptotic power (· · · · · ·) and empirical power (———) of the b2-star test for the distributions: a) wrapped exponential, b) wrapped half-normal, c) WSNC (λ=5), d) WSNC (λ=2). The six curves of each type correspond to sample sizes of 20, 30, 50, 100, 200 and 500, the power increasing with sample size. The dashed horizontal line delimits the nominal significance level of $\alpha = 0.05$.

We have also included in the plots making up these two figures the theoretical asymptotic power functions calculated using (5.5.1) and (5.5.2), respectively. The agreement between the empirical results and their theoretical counterparts for the b2-star test is excellent, even for sample sizes as small as 20. In general, the disparities between the two sets of results for the theta-bar test are more pronounced, the lack of agreement being greatest for small-sized samples drawn from close to isotropic parent populations. The overzealous inclination of the theta-bar test to reject symmetry under these conditions is in keeping with its non-conservatism established in Section 5.6.2.1.

Returning to the results portrayed in Figure 5.3, we see that the power of the b2-star test is greatest for middling values of ρ. The pronounced power deficiency of

the test for ρ-values of around 0.85 is consistent with the observations made in Section 5.3.1 regarding the expected value of b_2^* in this region of concentration. Notwithstanding the poor performance of the test for such ρ-values, as a comparison of the results in Figures 5.3 and 5.4 attests, for data distributed around a considerable arc of the unit circle the b2-star test is considerably more powerful than the theta-bar test. Moreover, the b2-star test does not suffer from the problems of non-conservatism displayed by the theta-bar test for data from close to uniform populations. However, for data drawn from more concentrated distributions one might feel inclined to employ the theta-bar test.

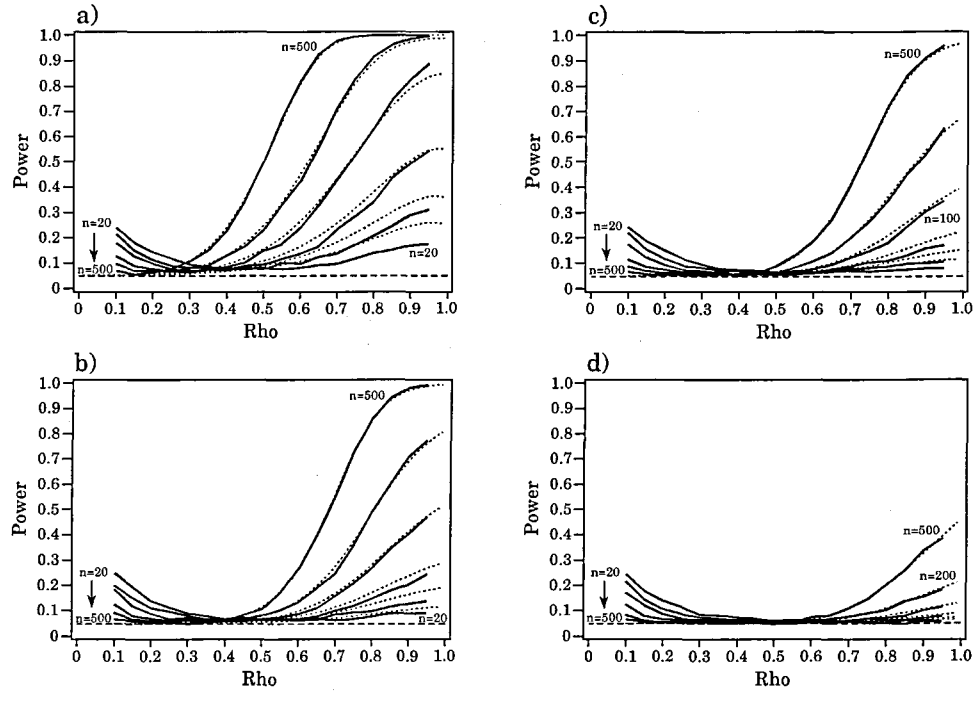

Figure 5.4 Theoretical asymptotic power (· · · · · ·) and empirical power (———) of the theta-bar test for the distributions: a) wrapped exponential, b) wrapped half-normal, c) WSNC (λ=5), d) WSNC (λ=2). The six curves of each type correspond to sample sizes of 20, 30, 50, 100, 200 and 500. The dashed horizontal line delimits the nominal significance level of $\alpha = 0.05$.

In fact, the theta-bar test is not the only test which outperforms the b2-star test when the parent population is more concentrated. As we might have expected, the circular analogues of the linear tests perform well under these conditions. In Figures 5.5-5.8 we portray some of the empirical results obtained for the b2-star test and the circular versions of the runs and modified runs tests. The latter two tests were found to consistently outperform the theta-bar and Wilcoxon tests. Each of the figures represents the results for a specific parent population. The low power

of all three tests exhibited in Figure 5.8 reflects the fact that the WSNC (λ=2) distribution is close to being symmetric, as can be appreciated from Figure 4.2d of Chapter 4. Also delimited in each plot is the ρ-value for which the empirical power of the b2-star test equals that of the circular analogue of the modified runs test. From a consideration of these four figures we see that the b2-star test consistently outperforms the other two tests for ρ-values less than, approximately, 0.45. Similarly, for values of ρ in excess of around 0.7, the modified runs test consistently outperforms, or is as powerful as, the other two tests.

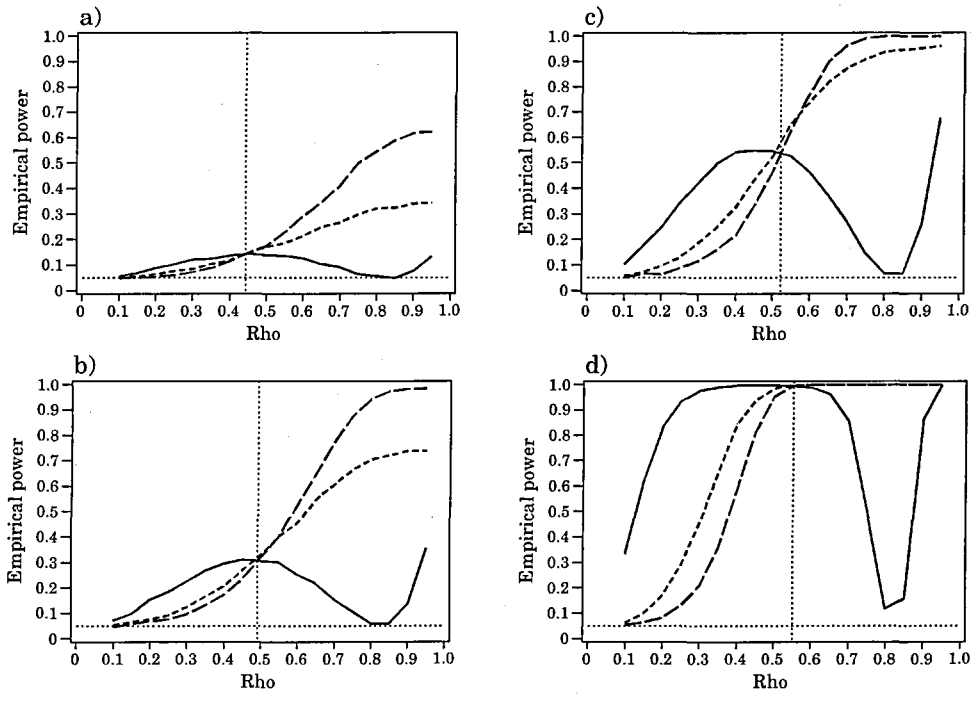

Figure 5.5 Empirical power against the wrapped exponential distribution for sample sizes of: a) 20, b) 50, c) 100, d) 500. The results represented are for the: b2-star (——), runs (– – –) and modified runs (— —) tests. The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$. The dotted vertical line indicates the point at which the empirical power of the b2-star test equals that of the modified runs test.

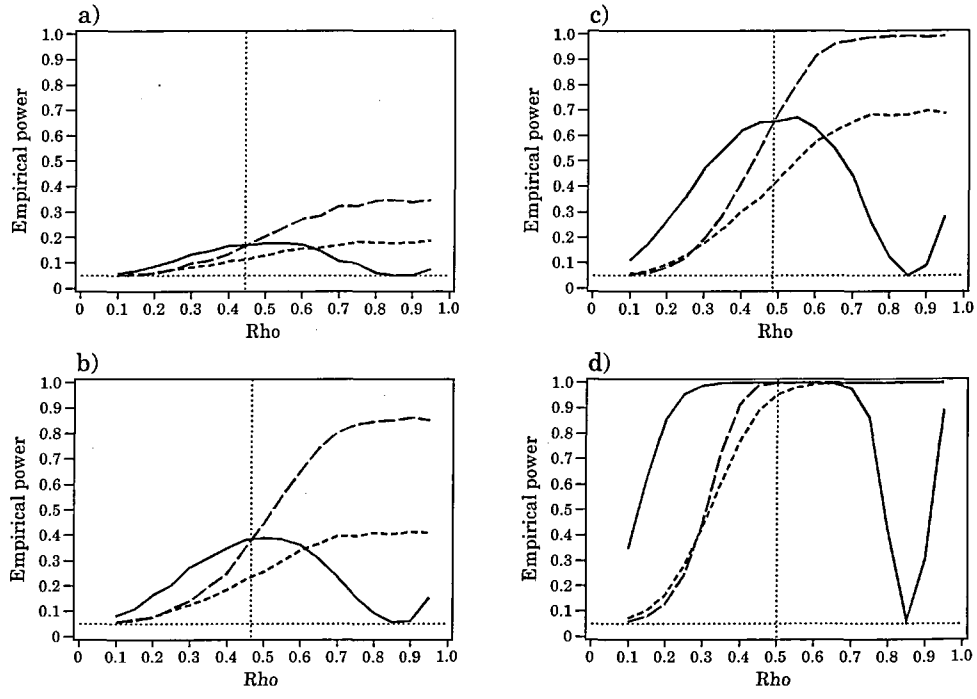

Figure 5.6 Empirical power against the wrapped half-normal distribution for sample sizes of: a) 20, b) 50, c) 100, d) 500. The results represented are for the: b2-star (——), runs (– – –) and modified runs (— —) tests. The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$. The dotted vertical line indicates the point at which the empirical power of the b2-star test equals that of the modified runs test.

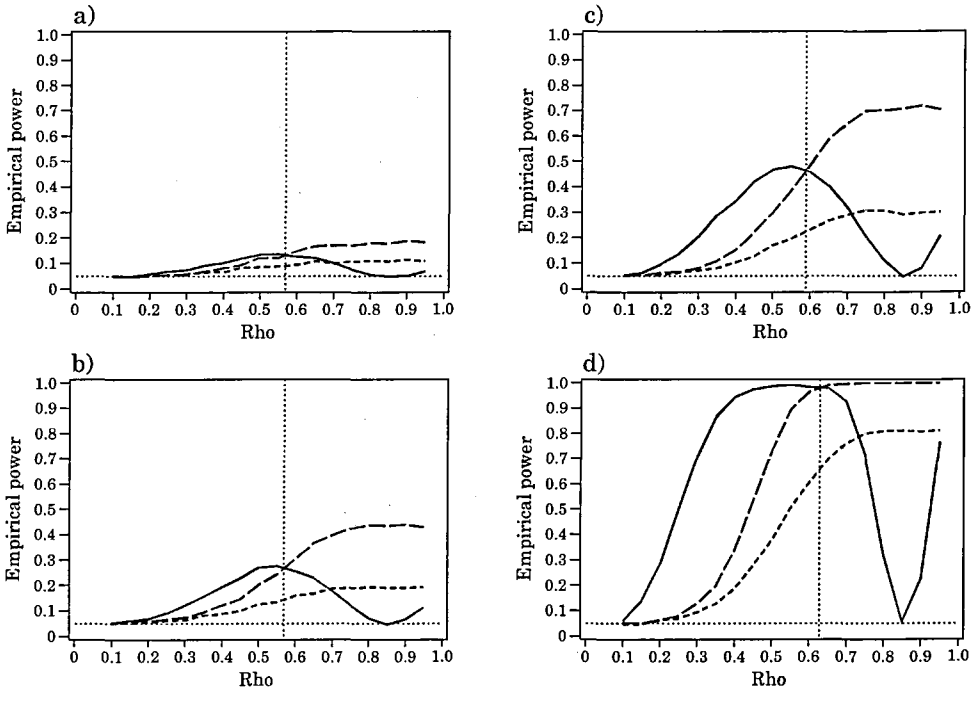

Figure 5.7 Empirical power against the WSNC (λ=5) distribution for sample sizes of: a) 20, b) 50, c) 100, d) 500. The results represented are for the: b2-star (——), runs (– – –) and modified runs (— —) tests. The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$. The dotted vertical line indicates the point at which the empirical power of the b2-star test equals that of the modified runs test.

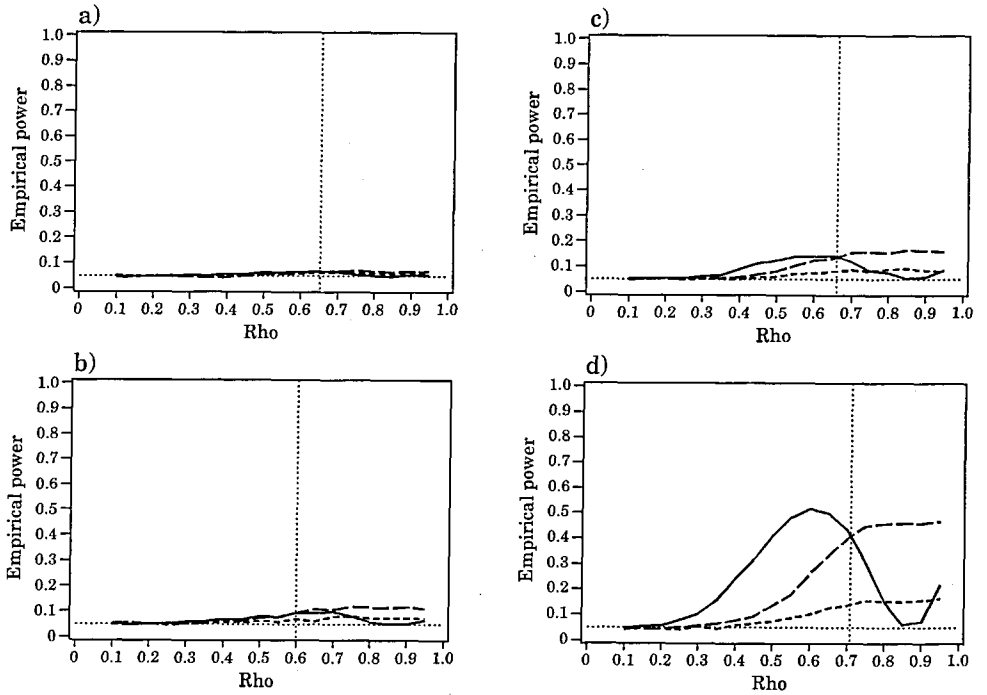

Figure 5.8 Empirical power against the WSNC (λ=2) distribution for sample sizes of: a) 20, b) 50, c) 100, d) 500. The results represented are for the: b2-star (——), runs (– – –) and modified runs (— —) tests. The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$. The dotted vertical line indicates the point at which the empirical power of the b2-star test equals that of the modified runs test.

5.6.3 A Simple Testing Strategy for Circular Data Drawn From Continuous Unimodal Populations

A more in-depth consideration of the full range of our results, including those for nominal significance levels of 10% and 1%, suggested the following testing strategy when the underlying distribution can be assumed to be continuous and unimodal.

> If the sample size is less than 100, apply the b2-star test if the sample mean resultant length, $\overline{R}$, is less than 0.45, and the circular analogue of the modified runs test if $\overline{R} > 0.65$. For a sample of 100 or more observations, apply the b2-star test if $\overline{R}$ is less than 0.5, and the circular analogue of the modified runs test if $\overline{R} > 0.7$. If in either case the value of $\overline{R}$ lies between the specified limits, it would be advisable to compare the results obtained from both tests to check for any glaring disparities.

Given the rather limited scope of the skew distributions used in our simulation study, the limits for $\overline{R}$ referred to in the above testing strategy should be interpreted as guidelines, not hard and fast restrictions. From a practical point of view, for instance, the quoted upper bounds on $\overline{R}$ for the optimality of the b2-star test may well be too low as they were established on the basis of the empirical power results for the wrapped exponential and wrapped half-normal distributions, two extremely skew models perhaps of limited relevance to the modelling of real circular data.

5.7 The Use of the New Procedures as Tests of Symmetry Against Rotation Alternatives

Having considered the power of our new procedures as tests of circular reflective symmetry about a known median axis against skew alternatives, in this section we explore their power as omnibus tests of circular reflective symmetry about a specified median axis against rotation alternatives. This is the testing scenario studied by Schach (1969) for which, as has already been mentioned, the circular analogues of the Wilcoxon and sign tests are usually employed as omnibus test procedures.

The circular analogue of the sign test rejects symmetry about the specified median axis in favour of symmetry about a displaced median axis if the number of the differences $\{\theta_i^* : i = 1,...,n\}$ in $[0,\pi)$ is either extremely large or extremely small when compared with the critical values of the binomial distribution with parameters n and 1/2.

Given the poor performance of the runs and modified runs tests for the equivalent testing context for linear data, one would not expect the circular analogues of these tests to have high power as tests of symmetry against rotation alternatives. Nevertheless, at this juncture we do not exclude them from our deliberations. On the other hand, given the measures upon which our two new procedures are based, one might expect them to provide powerful competitors to the circular analogues of the sign and Wilcoxon tests.

In Section 5.7.1 we derive the asymptotic power of our new procedures as tests of circular reflective symmetry about a specified median axis against rotation alternatives. The details of a simulation study designed to compare the small-sample power properties of the two new tests with those of the sign, Wilcoxon, runs and modified runs tests are given in Section 5.7.2.

5.7.1 The Asymptotic Power of the Two New Tests Against Rotation Alternatives

5.7.1.1 Power of the b2-star Test

Suppose under the null hypothesis, H_0, the underlying distribution is symmetric about a specified median direction, $\tilde{\mu}$, whilst under the alternative hypothesis, H_1, the underlying distribution is symmetric about a displaced median direction, $\tilde{\mu}+\delta$. Substituting these two hypotheses for those used in the original definition of the b2-star test given in Section 5.3.1, the implementation of the test is as described there.

Under H_0, let $\theta_{01},...,\theta_{0n}$ be n independently and identically distributed random variables from an angular distribution which is symmetric about $\tilde{\mu}$ and for which $\rho \neq 1$. Then, given the results in Section 5.3.1 and the fact that the distribution is symmetric about $\tilde{\mu}$, the asymptotic distribution of b_2^* under H_0 is normal with mean $E_0\left(b_2^*\right)=\beta_{2,0}^*$ and variance $\text{var}_0\left(b_2^*\right)=\frac{1}{2n}\left(1-\alpha_{4,0}^*\right)$, where $\alpha_{p,0}^* = E\left\{\cos p(\theta_{0i}-\tilde{\mu})\right\}$ and $\beta_{p,0}^* = E\left\{\sin p(\theta_{0i}-\tilde{\mu})\right\}=0$.

Under H_1, let $\theta_{11},...,\theta_{1n}$ be n independently and identically distributed random variables from an angular distribution which is symmetric about $\tilde{\mu}+\delta$ and for which $\rho \neq 1$. The asymptotic distribution of b_2^* under H_1 is normal with mean $E_1\left(b_2^*\right)$ and variance $\text{var}_1\left(b_2^*\right)$. We obtain,

$$\begin{aligned}
E_1\left(b_2^*\right) &= E\left\{\frac{1}{n}\sum_{i=1}^{n}\sin 2(\theta_{1i}-\tilde{\mu})\right\} \\
&= E\left\{\sin 2\left(\theta_{0i}+\delta-\tilde{\mu}\right)\right\} \\
&= E\left[\sin 2\left\{\left(\theta_{0i}-\tilde{\mu}\right)+\delta\right\}\right] \\
&= E\left\{\sin 2\left(\theta_{0i}-\tilde{\mu}\right)\cos 2\delta+\cos 2\left(\theta_{0i}-\tilde{\mu}\right)\sin 2\delta\right\} \\
&= \beta_{2,0}^*\cos 2\delta+\alpha_{2,0}^*\sin 2\delta \\
&= \alpha_{2,0}^*\sin 2\delta=\beta_{2,\delta}^*, \text{ say.}
\end{aligned}$$

Also,

$$\text{var}_1\left(b_2^*\right)=\frac{1}{2n}\left\{1-\alpha_{4,\delta}^*-2\left(\beta_{2,\delta}^*\right)^2\right\},$$

where

$$\begin{aligned}\alpha^*_{4,\delta} &= E\{\cos 4(\theta_{1i} - \tilde{\mu})\} \\ &= E[\cos 4\{(\theta_{0i} - \tilde{\mu}) + \delta\}] \\ &= E\{\cos 4(\theta_{0i} - \tilde{\mu})\cos 4\delta - \sin 4(\theta_{0i} - \tilde{\mu})\cos 4\delta\} \\ &= \alpha^*_{4,0}\cos 4\delta - \beta^*_{4,0}\sin 4\delta \\ &= \alpha^*_{4,0}\cos 4\delta .\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{var}_1(b_2^*) &= \frac{1}{2n}\left\{1 - \alpha^*_{4,0}\cos 4\delta - 2\sin^2 2\delta\,(\alpha^*_{2,0})^2\right\} \\ &= \frac{1}{2n}\left[1 - (\alpha^*_{2,0})^2 + \cos 4\delta\left\{(\alpha^*_{2,0})^2 - \alpha^*_{4,0}\right\}\right].\end{aligned}$$

Given these results, it follows that, for a significance level of $100\alpha\%$, the asymptotic power of the b2-star test against rotation alternatives is

$$1 - \Phi\left(\frac{z_{\alpha/2}\{\mathrm{var}_0(b_2^*)\}^{1/2} - E_1(b_2^*)}{\{\mathrm{var}_1(b_2^*)\}^{1/2}}\right) + \Phi\left(\frac{-z_{\alpha/2}\{\mathrm{var}_0(b_2^*)\}^{1/2} - E_1(b_2^*)}{\{\mathrm{var}_1(b_2^*)\}^{1/2}}\right). \quad (5.7.1)$$

The form taken by (5.7.1) as a function of ρ is exhibited in Figure 5.9 for the four symmetric distributions referred to previously in Section 5.6.1.

5.7.1.2 Power of the Theta-bar Test

When implementing the theta-bar test in this testing scenario, we assume the underlying distribution to be symmetric and to have a unique central direction. Under the null hypothesis, H_0, the underlying distribution is assumed to be symmetric about $\tilde{\mu} = \mu$, whilst under the alternative hypothesis, H_1, the underlying distribution is symmetric about the displaced central direction, $\tilde{\mu} + \delta = \mu + \delta$. Substituting these two hypotheses in place of their counterparts used in the original definition of the theta-bar test given in Section 5.3.2, one carries out the test as described there.

Given the results referred to in Section 5.3.2 and the fact that the distribution is assumed to be symmetric about $\tilde{\mu} = \mu$, the asymptotic distribution of $\bar{\theta}$ under H_0 is normal with mean $E_0(\bar{\theta}) = \tilde{\mu}$ and variance $\mathrm{var}_0(\bar{\theta}) = \dfrac{(1 - \alpha^*_{2,0})}{2n\rho^2}$.

Under H_1, the asymptotic distribution of $\bar{\theta}$ is normal with mean $E_1(\bar{\theta}) = \tilde{\mu} + \delta$ and variance $\mathrm{var}_1(\bar{\theta}) = \mathrm{var}_0(\bar{\theta})$. It therefore follows that, for a

significance level of $100\alpha\%$, the asymptotic power of the theta-bar test against rotation alternatives is

$$1-\Phi\left[z_{\alpha/2}-\frac{\delta}{\left\{\mathrm{var}_0\left(\bar{\theta}\right)\right\}^{1/2}}\right]+\Phi\left[-z_{\alpha/2}-\frac{\delta}{\left\{\mathrm{var}_0\left(\bar{\theta}\right)\right\}^{1/2}}\right]. \qquad (5.7.2)$$

We illustrate the form taken by (5.7.2) as a function of ρ in Figure 5.10.

5.7.2 Monte Carlo Investigation of the Small-sample Power of the Tests Against Rotation Alternatives

In Section 5.6.2.1 we presented the details of a simulation experiment designed to explore the ability of the two new procedures to maintain the nominal significance level under the null hypothesis of an underlying distribution which is symmetric about a known median axis. Those results clearly hold equally well for the present testing set-up in which the median axis is specified rather than known. In order to investigate the small-sample power of the new procedures as tests of symmetry against rotation alternatives, and to compare their power characteristics with those of the circular analogues of the sign, Wilcoxon, runs and modified runs tests, we conducted another simulation experiment.

Our study used data simulated from the same four symmetric unimodal distributions employed in the Monte Carlo experiment reported in Section 5.6. We explored the power of the various tests for the same sample sizes, ρ-values and nominal significance levels as used in that earlier study. For δ, the parameter representing the displacement in location, we used values of 2°, 4°, 6° and 8° from the zero direction. For each distribution, sample size, ρ and δ combination, 5000 samples were simulated. The sign, Wilcoxon and runs tests were randomized so as to have sizes as exactly specified.

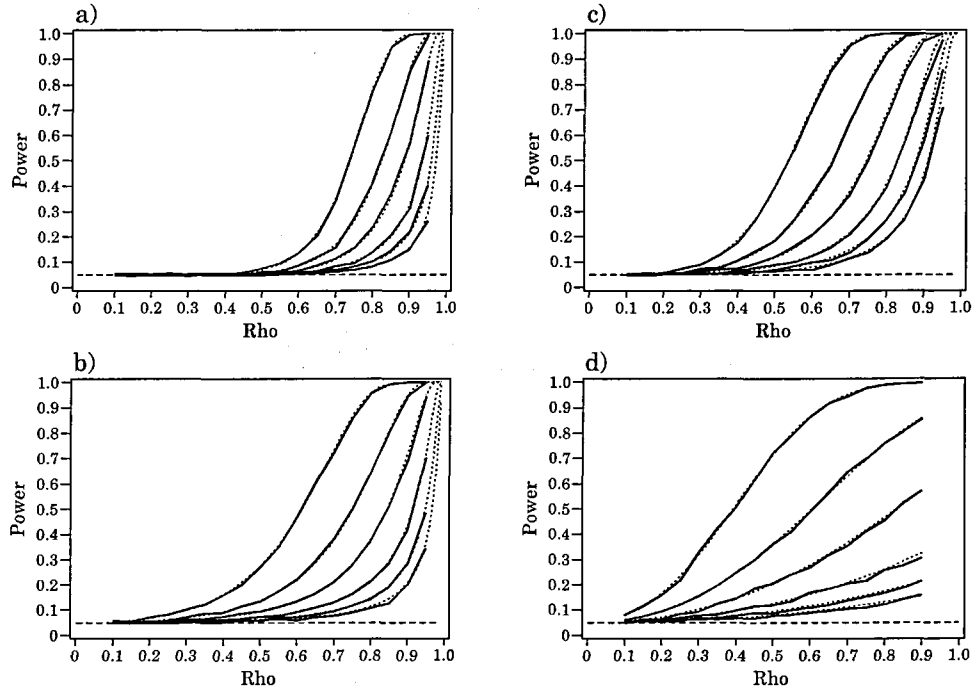

Figure 5.9 Theoretical asymptotic power (· · · · · ·) and empirical power (———) of the b2-star test against rotation alternatives for $\delta = 6°$ and the distributions: a) wrapped normal, b) wrapped Laplace, c) wrapped Cauchy, d) wrapped normal and uniform mixture. The six curves of each type correspond to sample sizes of 20, 30, 50, 100, 200 and 500, the power increasing with sample size. The dashed horizontal line delimits the nominal significance level of $\alpha = 0.05$.

For illustrative purposes, in Figures 5.9 and 5.10 we present the empirical results obtained for the b2-star and theta-bar tests, respectively, corresponding to a nominal significance level of 5% and a value of δ of $6°$. The standard error of any empirical result represented in these two figures is, at most, 0.007. Also included in these plots are the corresponding theoretical asymptotic power functions of the two tests calculated from (5.7.1) and (5.7.2), respectively. We note the generally close agreement between the two sets of curves, even for sample sizes as small as 20. Again, the major disparities between the theoretical and empirical results are those for the theta-bar test when used with data sampled from close to isotropic populations. The artificially high power of the theta-bar test under these conditions is consistent with our earlier findings portrayed in Figures 5.2 and 5.4.

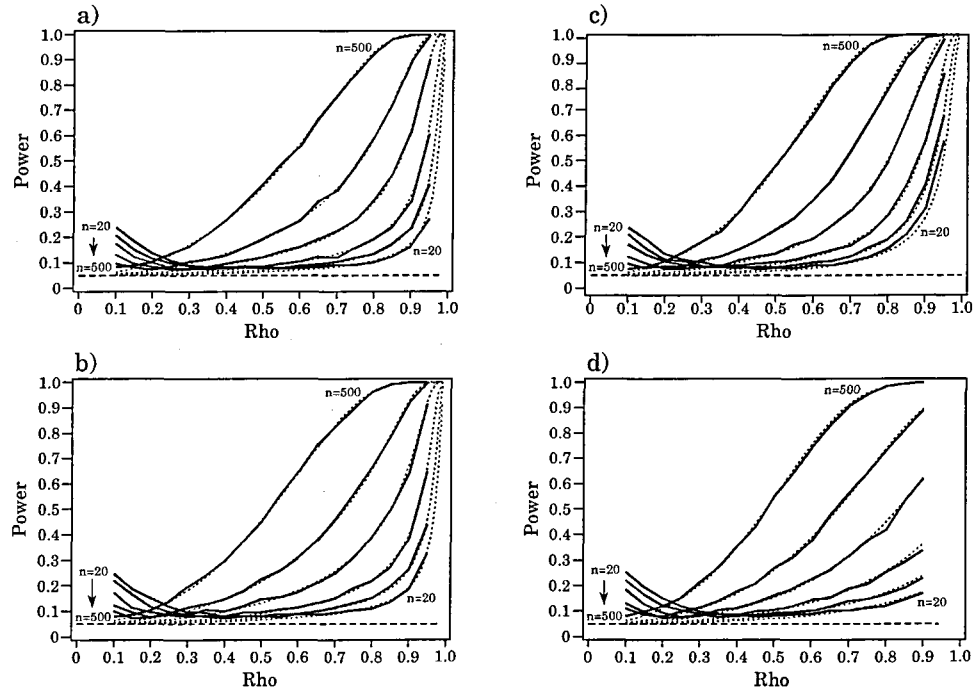

Figure 5.10 Theoretical asymptotic power (· · · · · ·) and empirical power (———) of the theta-bar test against rotation alternatives for $\delta = 6^\circ$ and the distributions: a) wrapped normal, b) wrapped Laplace, c) wrapped Cauchy, d) wrapped normal and uniform mixture. The six curves of each type correspond to sample sizes of 20, 30, 50, 100, 200 and 500. The dashed horizontal line delimits the nominal significance level of $\alpha = 0.05$.

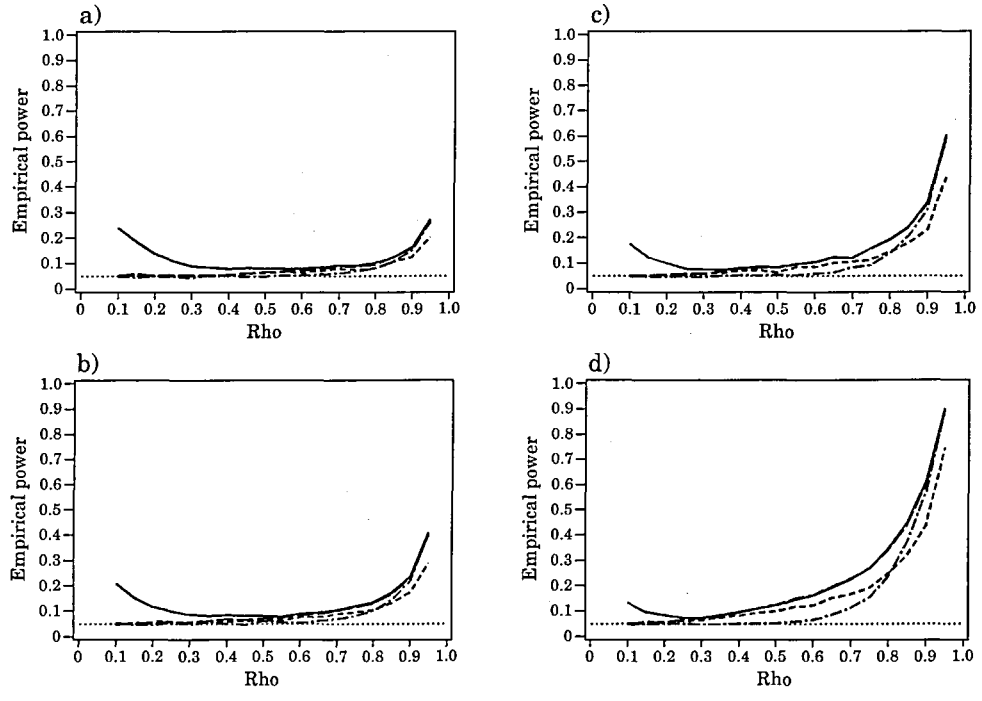

Figure 5.11 Empirical power against rotation alternatives for the wrapped normal distribution, $\delta = 6^\circ$ and sample sizes of: a) 20, b) 30, c) 50, d) 100. The results represented are for the: theta-bar (———), b2-star (— · —), sign (– – – –) and Wilcoxon (— — —) tests. The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$.

As expected, the power of the circular analogues of the runs and modified runs tests was found to be consistently lower than that of the other four tests. In Figures 5.11-5.14 we present representative results so as to aid a comparison of the power characteristics of those four tests. The empirical results portrayed correspond to sample sizes of 20, 30, 50 and 100, and a displacement in location under the alternative hypothesis of $\delta = 6°$.

From Figure 5.11 it can be seen that the theta-bar and Wilcoxon tests have almost identical power characteristics when the underlying distribution is wrapped normal, apart from when ρ is small. Both tests are more powerful than the sign and b2-star tests, but one would be ill-advised to use the theta-bar test for data drawn from highly dispersed cases of the distribution due to its pronounced non-conservatism under such conditions.

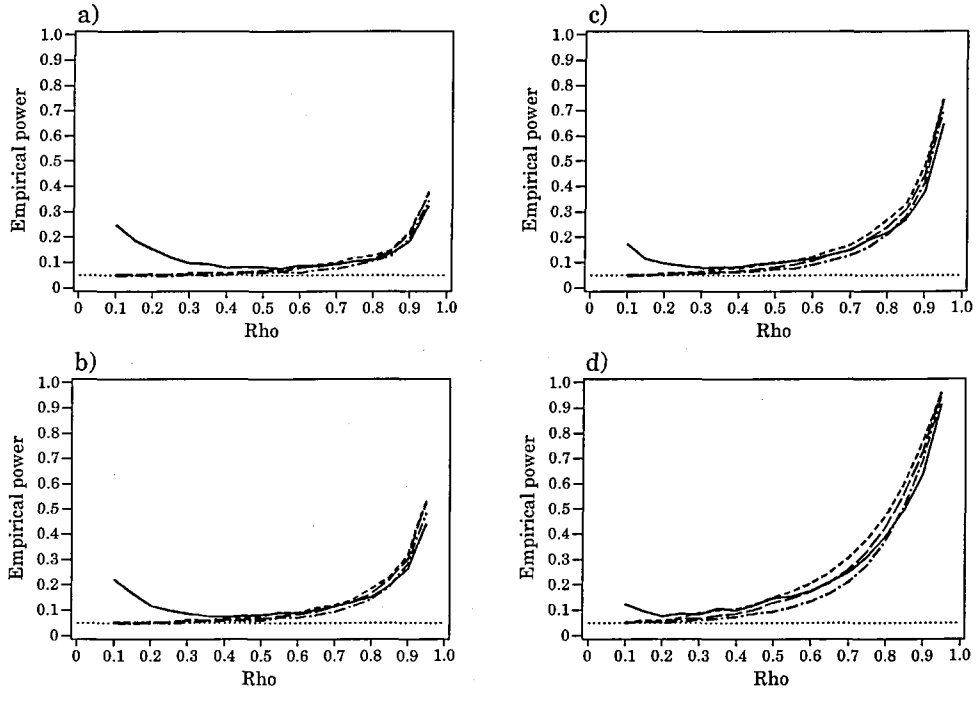

Figure 5.12 Empirical power against rotation alternatives for the wrapped Laplace distribution, $\delta = 6°$ and sample sizes of: a) 20, b) 30, c) 50, d) 100. The results represented are for the: theta-bar (———), b2-star (— · —), sign (– – – –) and Wilcoxon (— — —) tests. The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$.

Considering Figures 5.12 and 5.13, it can be observed that the sign test outperforms the other tests when the underlying population is wrapped Laplace or wrapped Cauchy. Interestingly, from Figure 5.14 we see that, for large n, the b2-

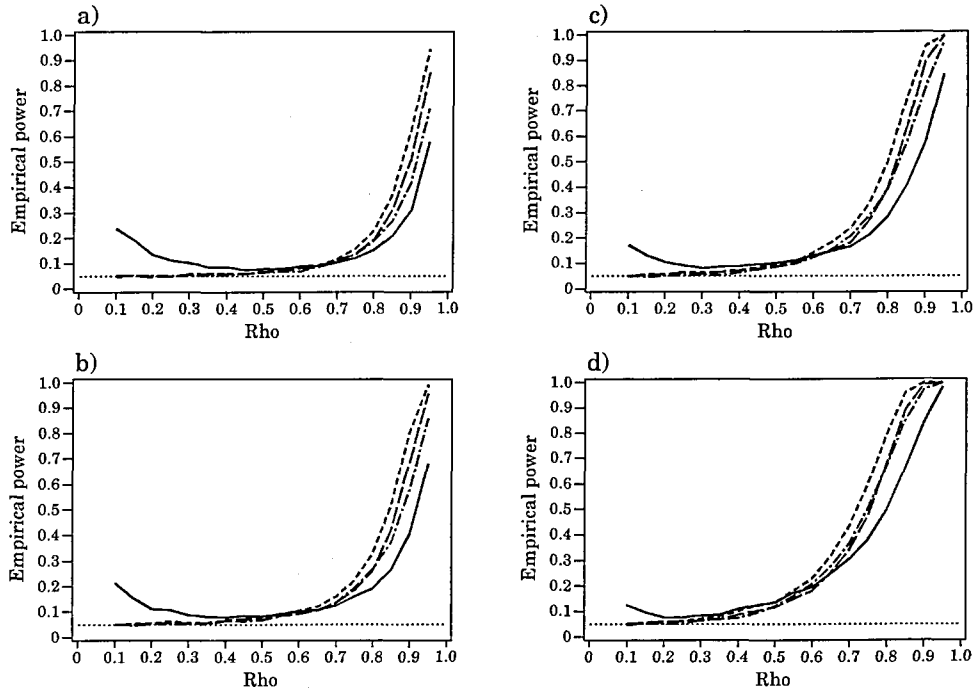

Figure 5.13 Empirical power against rotation alternatives for the wrapped Cauchy distribution, $\delta = 6°$ and sample sizes of: a) 20, b) 30, c) 50, d) 100. The results represented are for the: theta-bar (———), b2-star (— · —), sign (– – – –) and Wilcoxon (— — —) tests. The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$.

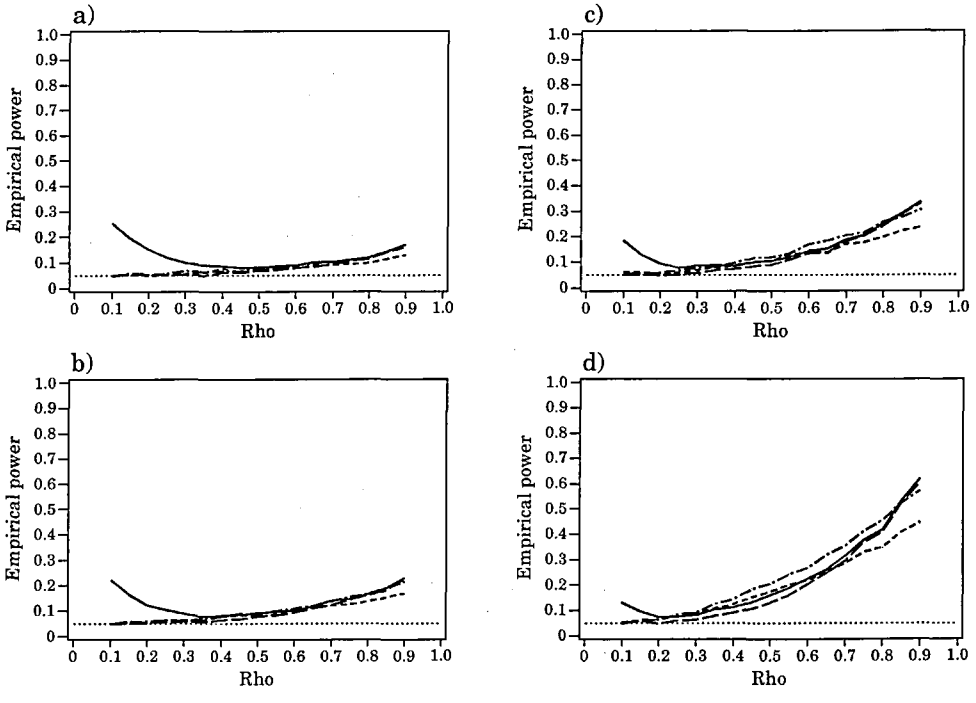

Figure 5.14 Empirical power against rotation alternatives for the wrapped normal and uniform mixture distribution, $\delta = 6°$ and sample sizes of: a) 20, b) 30, c) 50, d) 100. The results represented are for the: theta-bar (———), b2-star (— · —), sign (– – – –) and Wilcoxon (———) tests. The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$.

star test generally has the best power characteristics of the four tests considered when the parent population is the wrapped normal and uniform mixture distribution. However, as ρ approaches its maximum possible value for this distribution (≈ 0.9), with the distribution tending to a limiting normal distribution, the theta-bar and Wilcoxon tests dominate.

5.8 Use and Limitations of the Various Test Procedures

The strategy introduced in Section 5.6.3 for testing underlying unimodal distributions for symmetry about a known median axis against skew alternatives is based upon the use of the b2-star and the circular analogue of the modified runs tests. The assumptions underlying the b2-star test are minimal. Indeed, the only assumption made in addition to $\tilde{\mu}$ being a known median direction is that the underlying distribution is not a point distribution. In contrast, the use of the circular analogue of the modified runs test is more restricted as, strictly, the test requires the underlying distribution to be continuous. We also note that the b2-star test is a 'true' test of symmetry about a known median axis. By this we mean that the conclusions drawn from the test are unaffected if we replace $\tilde{\mu}$ by $\tilde{\mu}+\pi$. This, however, is not the case for the circular analogue of the modified runs test. Indeed, the modified runs test was specifically designed to be powerful against asymmetry in the tails of a linear distribution and therefore its circular analogue will not perform well for underlying distributions whose asymmetry manifests itself in regions close to the known median direction $\tilde{\mu}$.

Clearly, our developed strategy does not apply when the testing of circular reflective symmetry against rotation alternatives is of interest. In this context, the results presented in Section 5.7.2 highlight the difficulties inherent in identifying a consistently powerful omnibus test. Whilst the theta-bar test is competitive when the parent population is the wrapped normal, wrapped Laplace or wrapped mixture distribution, apart, of course, from when ρ is small, it does not perform well for an underlying wrapped Cauchy distribution. In a similar vein, the b2-star test performs best for the mixture distribution, moderately well for the wrapped Cauchy distribution, and not particularly well when the parent population is wrapped normal or wrapped Laplace. In addition to their power characteristics, the extreme simplicity of the sign test and the availability of software capable of performing the

Wilcoxon test certainly count greatly in favour of their usage. However, as our results attest, the blanket application of these two procedures can result in appreciable power loss. Moreover, the circular analogue of the Wilcoxon test is not a true test of symmetry about a specified median axis against rotation alternatives, in the sense that the conclusions reached upon using it will depend on whether $\tilde{\mu}$ or $\tilde{\mu}+\pi$ is used as the reference direction. Given our earlier comments, there is equally little to recommend the blanket usage of either the b2-star or the theta-bar tests in this testing scenario.

5.9 Illustrative Examples

In this section we apply the testing strategy developed in Section 5.6.3 to four data sets collected during animal orientation experiments.

5.9.1 Orientations of Red Ants

We first consider two data sets recorded during an orientation experiment with red wood ants (*Formica rufa* L.) described in Jandar (1957). Ants were placed singly in the centre of an arena and the initial direction in which they moved in relation to visual stimuli was recorded to the nearest $10°$. In a first version of the experiment (A) a black target was placed at an angle of $180°$ from the zero direction, while in a second (B), two black targets were used; one positioned at $90°$ from the zero direction, and the other at $180°$. The ants tended to move towards these targets. A basic question of interest in this context is whether the directions followed by the ants are distributed symmetrically about the central directions of $180°$ and $135°$, respectively, associated with the positions of the targets used in the two versions of the experiment.

The frequency distribution of the directions in which the 730 ants used in version A of the experiment moved is reproduced in Table 5.1 and represented graphically in the form of a raw circular plot in Figure 5.15. The sample median direction for these data is $180°$, identical to the direction associated with the position of the single black target. Hence it would indeed appear reasonable to assume the underlying median direction to be known and equal to $180°$. The sample mean resultant length is 0.620 and thus according to our developed testing

strategy we should compare the results obtained from applying both the b2-star and modified runs tests.

Table 5.1 Frequency distribution of the directions (in degrees) followed by the 730 ants in version A of an orientation experiment in which a single black target was positioned at 180° from the zero direction.

Angle	Frequency	Angle	Frequency	Angle	Frequency	Angle	Frequency
0	10	90	10	180	110	270	5
10	5	100	10	190	95	280	5
20	5	110	5	200	70	290	5
30	5	120	5	210	35	300	5
40	5	130	15	220	30	310	5
50	10	140	20	230	20	320	5
60	10	150	40	240	5	330	5
70	10	160	50	250	5	340	5
80	10	170	75	260	10	350	10

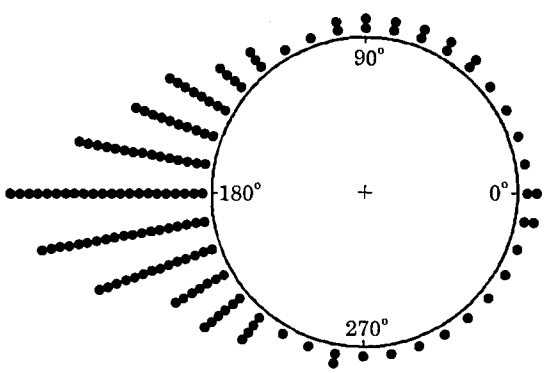

Figure 5.15 Raw circular plot of the ant orientation data of Table 5.1. Each dot represents the direction followed by five ants.

However, the data under consideration are discrete and so before being able to apply the modified runs test one first needs to employ some device in order to resolve the related problems of ties and the enumeration of runs. The device we used in order to achieve this was to add a random angle from the uniform distribution on $(-1\times10^{-8}, 1\times10^{-8})$ to each original angle measured in radians. Although this continuity inducing device resolves the problems arising from ties, it does so at the expense of introducing a p-value which is stochastic. Whilst we would not expect the conclusions reached from different randomizations of the original

data to differ greatly, such disparities are possible nonetheless. For moderate to large samples they are, however, unlikely.

The p-values obtained from applying the b2-star and modified runs tests with $\tilde{\mu} = 180°$ were 0.011 and 0.084, respectively. These two p-values thus provide evidence that the underlying distribution is, in fact, not symmetric about $\tilde{\mu} = 180°$. Given that the population median direction under the null hypothesis is identical to the sample median direction for these data, it would appear reasonable to interpret the rejection of underlying circular reflective symmetry as indicating that the parent population is asymmetric about $\tilde{\mu} = 180°$. It is therefore rather unfortunate that Batschelet (1981, pp. 48,49) chose to fit a von Mises distribution, and Sengupta & Pal (2001) a mixture distribution with symmetric wrapped stable and circular uniform components, to these data.

The frequency distribution presented in Table 5.2 summarizes the directions taken by the 1260 ants used in version B of the experiment. Figure 5.16 provides a graphical representation of the same data in the form of a raw circular plot. We note that the form taken by the sample distribution is perhaps what one might expect for an underlying mixture distribution with two equally weighted, and potentially symmetric, unimodal component distributions, one centred upon 90° and the other upon 180°.

Table 5.2 Frequency distribution of the directions (in degrees) followed by the 1260 ants in version B of an orientation experiment in which two black targets were positioned at 90° and 180° from the zero direction.

Angle	Frequency	Angle	Frequency	Angle	Frequency	Angle	Frequency
0	15	90	70	180	75	270	10
10	10	100	75	190	60	280	15
20	10	110	70	200	40	290	20
30	15	120	55	210	20	300	15
40	20	130	65	220	25	310	10
50	25	140	65	230	20	320	15
60	40	150	70	240	15	330	15
70	50	160	65	250	15	340	15
80	55	170	75	260	10	350	15

The sample median direction for these data is equal to the central direction defined by the experimental layout, i.e. 135°. The sample mean resultant length is 0.445, and so according to our testing strategy we should apply the b2-star test. The p-value for the b2-star test used with $\tilde{\mu} = 135°$ was found to be 0.818, and so we have no reason to reject the null hypothesis of circular reflective symmetry about $\tilde{\mu} = 135°$. For purely comparative purposes, the corresponding p-value for the circular analogue of the modified runs test, obtained after applying the continuity inducing device described in the analysis of the previous data set, was found to be 0.648. Again, this result provides us with no evidence to reject the null hypothesis.

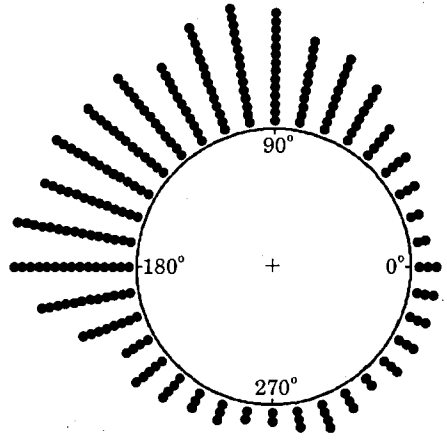

Figure 5.16 Raw circular plot of the ant orientation data of Table 5.2. Each dot represents the direction followed by five ants.

5.9.2 The Chinese Painted Quail Data Revisited

In Section 3.7.1 of Chapter 3 we presented data from an orientation experiment involving chinese painted quail reported by Merkel & Fischer-Klein (1973). The data were reproduced in Table 3.1 and represented graphically in Figure 3.1 of the same chapter. The question we address here is whether the underlying distribution is symmetric about $\tilde{\mu} = -15°$, the angle associated with the orientation of the last 0.5m stretch of the dog-leg corridor used in the design of the experiment. A visual inspection of Figure 3.1 would suggest that the underlying distribution is not symmetric. The sample median direction is −15.5°, and thus the assumption of a known median direction of $\tilde{\mu} = -15°$ would appear reasonable.

The sample mean resultant length for these data is 0.962 and so according to our testing strategy we should apply the circular analogue of the modified runs test. After applying the continuity inducing device referred to previously, the p-value for this test was found to be 0.089. Thus, the findings from the modified runs test partially support the initial visual impression of an underlying asymmetric distribution about $\tilde{\mu} = -15°$. Purely as a comparison, the p-values for the runs and b2-star tests were found to be 0.014 and 0.672, respectively. Thus, according to the runs test, the evidence against symmetry about $\tilde{\mu} = -15°$ is stronger than that implied by the result from the modified runs test. The p-value for the b2-star is consistent with the low power expected of this test for such a large mean resultant length and a sample size of 100.

5.9.3 Hisada's Dragonfly Data Revisited

In this last illustrative example we analyze once more the dragonfly data reported by Hisada (1972) which we presented in Section 3.7.3 of Chapter 3. Here we test the hypothesis that the underlying bimodal distribution is reflectively symmetric about a median axis assumed to be known and equal to the zero direction defined by the sun's azimuth. A visual inspection of Figure 3.5 suggests that symmetry about this direction is unlikely. The sample median direction is $50°$, but we note that the density of observations around the zero direction is extremely low.

Although our testing strategy was developed on the basis of findings for data drawn from unimodal distributions only, the sample mean resultant length for these data is 0.118 and such a low value of $\overline{R}$ would certainly favour the application of the b2-star test. In fact, the p-value for the b2-star test was found to be 0.020, providing strong evidence that the underlying distribution is not symmetric about the zero direction. Purely for comparative purposes, the p-values for the runs and modified runs tests, obtained after applying the continuity inducing device described earlier, were 0.050 and 0.142, respectively. We note that the results for the b2-star and runs tests tend to confirm our earlier findings in Section 3.7.3 regarding the asymmetry of the parent population from which these data were drawn.

5.10 Summary and Directions for Future Research

In this last section of the chapter we provide a summary of the content of the preceding sections and discuss potential lines of related future research.

5.10.1 Summary

The main focus of the present chapter has been the testing of circular reflective symmetry about a known median axis against skew alternatives. In Section 5.2 we provided a review of the background to this testing problem, relating it to the testing of linear data for symmetry about a known or specified median.

In Section 5.3 we introduced two new asymptotically distribution-free procedures, the b2-star and theta-bar tests, which can be used as omnibus tests of symmetry about a known median axis against skew alternatives. Circular analogues of three linear tests were described in Section 5.4, these tests being proposed as competitors to the b2-star and theta-bar tests.

In Section 5.5 we derived theoretical results for the asymptotic power of the b2-star and theta-bar tests as procedures for testing for circular reflective symmetry about a known median axis against skew alternatives. The details of a simulation study designed to explore and compare the operating characteristics of the various procedures were presented in Section 5.6. From a consideration of the results from this study, summarized in Section 5.6.2, we found that the b2-star, Wilcoxon, runs and modified runs tests maintained the nominal significance level well. However, the theta-bar test proved to be non-conservative for data sampled from highly dispersed parent populations, particularly when the sample size was small.

In Section 5.6.2.2 we considered the results for the power characteristics of the various procedures against skew alternatives. The b2-star test was identified as being a powerful test for data distributed over a considerable arc of the unit circle, but to be power deficient for data sampled from more concentrated parent populations. Given the theta-bar test's previously established non-conservatism, we concluded that its use as a test of symmetry about a known median axis against skew alternatives could not be recommended for data drawn from highly dispersed populations. Nevertheless, the power characteristics of this test for data distributed over a reduced arc of the unit circle were identified as being superior to those of the b2-star test. Our results also showed that the modified runs test had even better power properties than the theta-bar test for data drawn from continuous unimodal distributions with levels of concentration which are moderate to high. On the basis

of a comparison of the empirical power properties of the b2-star and modified runs tests, in Section 5.6.3 we proposed a simple testing strategy for circular data drawn from continuous unimodal populations.

Although the b2-star and theta-bar tests were originally conceived as tests of circular reflective symmetry about a known median axis against skew alternatives, in Section 5.7 we considered their use as tests of symmetry about a specified median axis against rotation alternatives. In Section 5.7.1 we derived the asymptotic power of the tests, and in Section 5.7.2 presented results from a simulation study conducted in order to explore and compare the small-sample power properties of the b2-star, theta-bar, sign, Wilcoxon, runs and modified runs tests in this testing set-up. The runs and modified runs tests were identified as having poor power properties, whilst the b2-star and theta-bar tests proved to be competitive with the sign and Wilcoxon tests for certain parent population and concentration combinations.

In Section 5.8 we discussed the usage and limitations of the various test procedures. The b2-star test was identified as relying on minimal assumptions and being a true test of circular reflective symmetry about a given median axis. The sign and runs tests are also true tests of circular reflective symmetry about a given median axis, but the Wilcoxon and modified runs tests are not. In addition, the use of the latter tests is more restricted as both tests require the parent population to be continuous. Use of the theta-bar test requires the assumption that $\rho \neq 0$, a constraint which excludes the uniform distribution and, at first sight, all multimodal distributions which are either cyclically symmetric, or have more than one axis of symmetry, or both. However, if the underlying distribution is thought to be multimodal with $\rho = 0$, and the number of modes, m, of the distribution can be established beforehand, one can always administer the theta-bar test after first applying the device of m-fold wrapping of the circle onto itself, as described in Section 4.3 of Chapter 4.

In Section 5.9 we illustrated the application of our testing strategy in the analysis of four data sets collected during animal orientation experiments. Our illustrative examples highlighted the problems inherent in the application of the modified runs test to real circular data, some form of continuity inducing device having to be applied prior to carrying out the test.

5.10.2 Directions for Future Research

In this chapter we introduced two new test procedures specifically designed for testing truly circular data for symmetry about a known median axis, $\tilde{\mu}$, against skew alternatives. Both tests are based on natural measures of skewness about $\tilde{\mu}$. As has been pointed out in Section 5.3.1, the b2-star test is an adaptation of the test based on $\bar{b}_2$ considered in Chapter 4. This pair of tests can be considered as the circular equivalent of a pair of tests for linear data considered by Gupta (1967); one based on the coefficient of skewness $g_1 = m_3 / m_2^{3/2}$ and the other on an adaptation of g_1 in which the sample mean in its definition is replaced by the population median. The second of our tests, based on the difference between the sample mean direction and $\tilde{\mu}$, effectively combines ideas for linear measures of skewness proposed early in the literature by Yule (1911) and Hotelling and Solomon (referred to by Kendall & Stuart (1963, p. 93)) and adapts them to the equivalent testing scenario for circular data. Also considered in our investigations were the circular analogues of some of the most powerful procedures known for testing linear data for symmetry against skew alternatives. Indeed, the testing strategy outlined in Section 5.6.3 reflects the strengths of a test designed specifically for truly circular data (the b2-star test) as well as those of the circular analogue of a powerful linear test (the modified runs test). For reasons given in Section 5.2, we did not consider the circular analogue of the hybrid test of Modarres & Gastwirth (1998) in our investigations. If the problems with the specification of the linear test can be resolved, it would be interesting to compare the operating characteristics of its circular analogue with those of the b2-star test and the circular analogue of the modified runs test. However, we note that, like the circular analogue of the modified runs test, the circular analogue of the hybrid test of Modarres & Gastwirth (1998), being based on a percentile modified two-sample Wilcoxon test, would not be a true test of circular reflective symmetry about a median axis in the sense described in Section 5.8. As with the circular analogue of the modified runs test, in theory its use would also be restricted to data drawn from continuous populations only.

Our testing strategy was developed on the basis of empirical results from a rather limited range of unimodal circular distributions. It would clearly be of interest to know just how robust this testing strategy is to the form of the underlying distribution, not only for other unimodal distributions but also for

multimodal distributions. Regarding this issue, we note that the published recommendations concerning the use of the runs test of Cohen & Menjoge (1988) and McWilliams (1990), the modified runs test of Modarres & Gastwirth (1996), the conditional test of Tajuddin (1994) and the hybrid test of Modarres & Gastwirth (1998), were established solely on the basis of empirical results obtained from Monte Carlo experiments involving data simulated from unimodal populations only.

Chapter 6 The Wrapped Skew-normal Distribution on the Circle

6.1 Introduction

In the preceding two chapters we considered test procedures which can be employed in the analysis of circular data in order to establish whether the assumption of underlying circular reflective symmetry is reasonable or not. But what if symmetry is rejected? What options are available to us in terms of modelling circular data which are skew? In Section 6.2 we provide a necessarily brief review of the asymmetric models for circular data which have been proposed in the literature. The remainder of the chapter is devoted almost exclusively to the wrapped skew-normal distribution on the circle, henceforth the WSNC distribution, a new model motivated by the need for distributional forms capable of modelling the asymmetry often manifested by real circular data.

In Section 6.3 we define the distribution and obtain its fundamental properties. We note that the derivation of the characteristic function which appears in Section 6.3.2 is due to Professor Toby Lewis.

The following four sections address issues of inference for the WSNC distribution. Sections 6.4, 6.5 and 6.6 discuss the estimation of the distribution's parameters. A viable approach to method of moments estimation is presented in Section 6.4, and maximum likelihood estimation is discussed in Section 6.5. A detailed evaluation of these competing approaches to estimation is given in Section 6.6. In Section 6.7 we consider procedures for testing for three limiting cases of the WSNC class.

The application of the developed methodology is illustrated in Section 6.8 using a large data set consisting of the headings of migrating birds. The chapter closes, in Section 6.9, with a summary of its content and a discussion of potential lines of related future research.

In the main, the treatment of the WSNC distribution given here follows that published in Pewsey (2000b). However, much of the content of Sections 6.2, 6.3, 6.4 and 6.7, and all of that of Sections 6.5 and 6.6, is new. In addition, the analysis of

the illustrative example presented in Section 6.7 differs in important ways from that given in Pewsey (2000b), and is effectively that presented at the 19th Leeds Annual Statistics Research Workshop which took place at the University of Leeds in July 2000, and at the 25th National Conference of Statistics and Operational Research held at the University of Jaén, Úbeda, Spain in November 2001.

6.2 Asymmetric Models for Circular Data

Of the continuous models for circular data, the best known and most frequently applied are unimodal and symmetric. Amongst this class of models figure the cardioid, wrapped Cauchy, wrapped normal and von Mises distributions. As was previously pointed out in Section 4.2 of Chapter 4, historically the von Mises distribution has played a dominating role in the analysis of circular data.

In contrast to the situation for linear data, relatively few asymmetric models have been proposed in the literature for circular data. Mardia (1972, pp. 51-53) considered a family of skew triangular distributions and the class of projected normal distributions. The latter class can be used to model bimodality as well as asymmetry.

Batschelet (1981, Section 15.6) refers to two distributional families capable of modelling skewness. The first is due to Papakonstantinou (1979) who proposed the class defined by the density

$$f(\theta)=\frac{1}{2\pi}+\frac{\kappa}{2\pi}\sin(\theta+\nu\sin\theta), \tag{6.2.1}$$

where $\theta\in[0,2\pi)$, $|\nu|<1$ and $|\kappa|\leq 1$. The shape of the distribution depends on the values taken by both κ and ν, the latter of the two parameters determining the skewness of the distribution. The special case of $\nu=0$ corresponds to the symmetric cosine distribution (see, for example, Batschelet (1981, p. 283)), whilst the uniform distribution results if $\kappa=0$.

The second class is a particular extension of the von Mises distribution, defined by the density

$$f(\theta)=c\exp\{\kappa\cos(\theta+\nu\cos\theta)\}, \tag{6.2.2}$$

where, once more, $|\nu|<1$, and c is a normalizing constant which must be calculated numerically. Again, ν plays the role of a skewness parameter, the von Mises distribution with concentration parameter κ resulting when $\nu=0$.

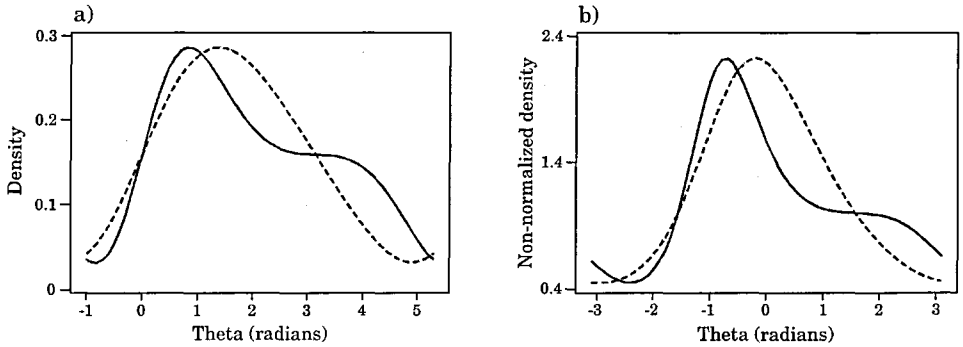

Figure 6.1 Linear plots of Equations: a) (6.2.1), b) (6.2.2). Both pairs of curves correspond to the choices $\kappa = 0.8$ and $\nu = 0.2$ (dashed curve) and $\nu = 0.98$ (solid curve). In b) the normalizing constant c has been set equal to 1.

Papakonstantinou (1979, pp. 28-29) derived the trigonometric moments of the density given in (6.2.1) and hence the second central sine moment of the distribution. Otherwise, we are unaware of any other published work extending the results of these two authors. Here we briefly mention two rather undesirable features of both families. Firstly, as defined, neither model contains a location parameter. Secondly, whilst the densities described by (6.2.1) and (6.2.2) are monotone decreasing about a main mode when $|\nu| < 1$, as $|\nu| \to 1$ they take on rather unappealing shapes. To illustrate this common behaviour, in Figure 6.1 we graph (6.2.1) and (6.2.2), the latter with $c = 1$, for $\kappa = 0.8$ and ν equal to 0.2 and 0.98. These changes in shape are a consequence of the fact that the parameter ν not only controls the skewness of the two distributions but, if allowed to freely vary, also determines the number of modes manifested by them. The constraint that $|\nu|$ should be less than 1 ensures that the derivatives of (6.2.1) and (6.2.2) do not vanish, but is not sufficient to avoid the changes in shape of both densities evident in Figure 6.1.

A simple yet highly useful means of generating skew models for circular data is to wrap a suitably chosen skew linear class onto the unit circle. This idea is by no means new, indeed its first use dates back at least to the work of Schmidt (1917). Mardia (1971) referred briefly to the possibility of wrapping the flexible Pearson family of distributions onto the circle, but to our knowledge this option has not subsequently been pursued in the literature. Another family of distributions of this type which has great potential in the modelling of skew circular data is the wrapped stable class. This family includes the symmetric wrapped normal and

wrapped Cauchy distributions as special cases. Relevant published work regarding this class is that of Winter (1947) and Sengupta & Pal (2001). The wrapped exponential distribution introduced in Section 4.4.2.2.1 of Chapter 4 provides another example of the application of this general approach to obtaining circular distributions from linear ones.

One of the most striking features of the treatment given in the literature to asymmetric models is its almost universal statistical superficiality. Whilst the basic properties of the various distributions are, in the main, well documented, it is hard, if not impossible, to find anything published regarding inference for such distributions. It is not, therefore, surprising to find that, as Fisher (1994, p. 56) comments, models capable of modelling asymmetry have rarely been applied in the analysis of circular data. In an attempt to redress this situation, we dedicate the greater part of this chapter to a consideration of the wrapped skew-normal distribution on the circle as a potential model for skew circular data.

6.3 Definition and Fundamental Properties of the WSNC Distribution

6.3.1 Definition and Limiting Cases

Following the notation of Chapter 1, suppose $Y_D \sim \mathrm{SN}_D(\xi,\eta,\lambda)$ with $-\infty<\xi<\infty$, $\eta>0$ and $-\infty<\lambda<\infty$. Wrapping the random variable Y_D onto the unit circle, we denote the resulting circular random variable as $\theta_D = Y_D \pmod{2\pi}$. Then θ_D has density

$$f(\theta;\xi,\eta,\lambda)=\frac{2}{\eta}\sum_{r=-\infty}^{\infty}\phi\left(\frac{\theta+2\pi r-\xi}{\eta}\right)\Phi\left\{\lambda\left(\frac{\theta+2\pi r-\xi}{\eta}\right)\right\}, \qquad (6.3.1)$$

for $0\le\theta<2\pi$. We will say that θ_D has a WSNC distribution with direct parameters (ξ,η,λ), and denote this distributional relation as $\theta_D \sim \mathrm{WSNC}_D(\xi,\eta,\lambda)$.

There are five limiting cases of the WSNC distribution. As $\eta\to 0$, the distribution tends to a point distribution, whereas, as $\eta\to\infty$, the limiting distribution is the circular uniform. When $\lambda=0$ the density (6.3.1) reduces to that of the wrapped normal distribution. Considering the two extremes of the λ scale, the distribution tends to the wrapped half-normal distribution and the wrapped

negative half-normal distribution as λ tends to $+\infty$ and $-\infty$, respectively. Circular plots of four $\mathrm{WSNC_D}(0,1,\lambda)$ densities with pertinent positive values of λ are presented in Figure 6.2.

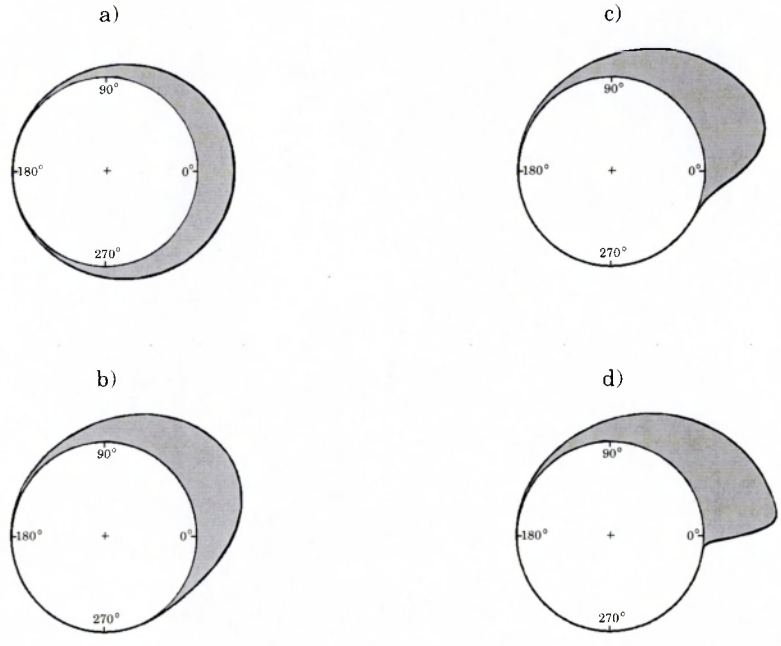

Figure 6.2 Circular plots of $\mathrm{WSNC_D}(0, 1, \lambda)$ densities with: a) $\lambda = 0$ (wrapped standard normal distribution); b) $\lambda = 2$; c) $\lambda = 5$; d) $\lambda = 20$.

6.3.2 Characteristic Function and Trigonometric Moments

In Section 4.4.2.2.1 we gave results for the median direction of the WSNC distribution. Here, we derive the characteristic function of the distribution and identify its trigonometric moments.

Consider the linear standard skew-normal random variable $X \sim \mathrm{SN}(\lambda)$. The characteristic function of X, $\psi_X(t)$, is given by

$$\psi_X(t) = E\left(e^{itX}\right) = E(\cos tX) + iE(\sin tX)$$

$$= 2\int_{-\infty}^{\infty} \cos tx\, \phi(x)\Phi(\lambda x)dx + i\,2\int_{-\infty}^{\infty} \sin tx\, \phi(x)\Phi(\lambda x)dx$$

which, using an obvious notation, we will represent as $\psi_X(t) = C(t,\lambda) + i\,S(t,\lambda)$. Now,

$$\frac{\partial C(t,\lambda)}{\partial \lambda} = 2\int_{-\infty}^{\infty} \cos tx\, \phi(x) x\, \phi(\lambda x) dx = 0,$$

the integrand being an odd function. Thus,

$$C(t,\lambda) = C(t,0) = \int_{-\infty}^{\infty} \cos tx\, \phi(x) dx = e^{-\frac{1}{2}t^2}.$$

Also,

$$\begin{aligned}
\frac{\partial S(t,\lambda)}{\partial t} &= 2\int_{-\infty}^{\infty} x \cos tx\, \phi(x) \Phi(\lambda x) dx \\
&= -2\int_{-\infty}^{\infty} \cos tx\, \Phi(\lambda x) d\phi(x) \\
&= 2\int_{-\infty}^{\infty} \phi(x)\{-t \sin tx\, \Phi(\lambda x) + \lambda \cos tx\, \phi(\lambda x)\} dx \\
&= -tS(t,\lambda) + 2\lambda \int_{-\infty}^{\infty} \cos tx \frac{1}{2\pi} \exp\{-\tfrac{1}{2}(1+\lambda^2)x^2\} dx.
\end{aligned}$$

So,

$$\begin{aligned}
\frac{\partial S(t,\lambda)}{\partial t} + tS(t,\lambda) &= (2/\pi)^{1/2} \frac{\lambda}{(1+\lambda^2)^{1/2}} \exp\{-\tfrac{1}{2}t^2/(1+\lambda^2)\} \\
&= b\delta \exp\{-\tfrac{1}{2}t^2/(1+\lambda^2)\}.
\end{aligned}$$

Thus,

$$\frac{\partial\{S(t,\lambda)e^{\frac{1}{2}t^2}\}}{\partial t} = b\delta\, e^{\frac{1}{2}\delta^2 t^2}$$

and hence

$$S(t,\lambda)e^{\frac{1}{2}t^2} = c + \int_0^{\delta t} be^{\frac{1}{2}v^2} dv.$$

Now, as $S(0,\lambda) = 0, c = 0$, and therefore,

$$S(t,\lambda) = e^{-\frac{1}{2}t^2} \int_0^{\delta t} be^{\frac{1}{2}v^2} dv.$$

Thus, the characteristic function of X is

$$\psi_X(t) = e^{-\frac{1}{2}t^2}\left(1 + i\int_0^{\delta t} be^{\frac{1}{2}v^2} dv\right).$$

It follows that the characteristic function of the general skew-normal random variable $Y_D = \xi + \eta X \sim \mathrm{SN}_D(\xi,\eta,\lambda)$, $\psi_{Y_D}(t)$, is given by

$$\begin{aligned}\psi_{Y_D}(t) &= e^{i\xi t}\psi_X(\eta t)\\ &= e^{i\xi t-\frac{1}{2}\eta^2 t^2}\left(1+i\int_0^{\delta\eta t} be^{\frac{1}{2}v^2}dv\right)\\ &= e^{i\xi t-\frac{1}{2}\eta^2 t^2}\{1+i\,\Im(\delta\eta t)\}\end{aligned}$$

where, for positive x,

$$\Im(x)=\int_0^x be^{\frac{1}{2}u^2}du,$$

and $\Im(-x)=-\Im(x)$. As we shall see, the function $\Im(\cdot)$ is highly important in terms of inference for the WSNC distribution. We used numerical methods of integration to calculate values of $\Im(x)$, although other approaches are possible as $\Im(x)$ is closely related to Dawson's integral (see, for example, Spanier & Oldham (1987, Chapter 42)). Table 6.1 provides some representative values of the function $\Im(x)$ for a range of x-values. These, along with all other required values of $\Im(x)$, were calculated using 15-point Gauss-Legendre quadrature (see, for example, Carnahan *et al.* (1969, Chapter 2)).

Table 6.1 Values of $\Im(x)$ for a range of values of x.

x	$\Im(x)$	x	$\Im(x)$
0.0	0.00000	6.0	8.99783×10^{6}
0.5	0.41621	8.0	8.00475×10^{12}
1.0	0.95344	10.0	4.17947×10^{20}
2.0	3.77312	15.0	3.85416×10^{47}
3.0	28.2385	20.0	2.89001×10^{85}
4.0	643.127	25.0	1.66619×10^{134}
5.0	44800.0	32.0	5.70151×10^{220}

Having derived the characteristic function of Y_D, that of $\theta_D = Y_D \pmod{2\pi}$, $\{\psi_{\theta_D,p} : p=0,\pm1,...\}$, is given by (see, for example, Mardia & Jupp (1999, Section 3.5.7)),

$$\psi_{\theta_D,p}=\psi_{Y_D}(p)=e^{ip\xi-\frac{1}{2}p^2\eta^2}\{1+i\,\Im(\delta\eta p)\}. \tag{6.3.2}$$

From (6.3.2), one obtains the cosine moments

$$\alpha_p = E(\cos p\theta_D)=\omega^{p^2}(\cos p\xi-\Im(\delta\eta p)\sin p\xi), \tag{6.3.3}$$

and the sine moments,

$$\beta_p = E(\sin p\theta_D) = \omega^{p^2}(\sin p\xi + \Im(\delta\eta p)\cos p\xi), \tag{6.3.4}$$

where $\omega = e^{-\frac{1}{2}\eta^2}$. Given these trigonometric moments, an alternative representation for the density of θ_D is,

$$f(\theta;\xi,\eta,\lambda) = \frac{1}{2\pi}\left[1 + 2\sum_{p=1}^{\infty}\omega^{p^2}\{\cos p(\theta-\xi) + \Im(\delta\eta p)\sin p(\theta-\xi)\}\right].$$

6.4 Method of Moments Estimation

6.4.1 Introduction

Having outlined the fundamental properties of the WSNC distribution, in this and the following two sections we consider the issue of estimating the distribution's parameters. As in previous chapters, we start with a discussion of method of moments estimation before moving on to a consideration of likelihood based inference.

Using the trigonometric moments in (6.3.3) and (6.3.4), one can, in principle, obtain method of moments (MM) estimates for the direct parameters of the WSNC distribution by equating the expressions for α_1, β_1 and β_2, say, with their sample counterparts a_1, b_1 and b_2, and solving for the direct parameters ξ, η and λ. Nevertheless, in addition to the problem as to exactly which of the trigonometric moments should be used in this exercise, there are many potential ways of solving the resulting systems of equations. We explored numerous variants of this approach but found in each case that the systems of equations generated by them sometimes had no solution. After trying in vain to identify a version of this approach which always led to a soluble system of equations, we were led to conclude that the problem was a consequence of the inherent instability of the estimates obtained when attempting to solve directly for ξ, η and λ.

6.4.2 Estimation Based on a Circular Parametrization

In the search for a set of parameters more amenable to moment based estimation, we pursued the logic of Azzalini's centred parametrization discussed in Section 1.4.2 of Chapter 1. Clearly, the centred parameters of the linear skew-normal distribution will not generally represent useful properties of a circular distribution. However, as we saw in Chapter 3, analogous measures which represent the central

location, dispersion and skewness of a circular distribution are at our disposal. Given the results derived in Chapter 3, we chose to work with the parametrization $(\mu_C, \rho, \bar{\beta}_2)$; $\mu_C \in [0, 2\pi)$ being the mean direction, $\rho \in [0, 1]$ the mean resultant length and $\bar{\beta}_2 \in [-1, 1]$ the second central sine moment. Here and in the remainder of the chapter we use μ with the subindex C to indicate that the parameter concerned is a circular one. To avoid potential confusion, later we use μ_L to denote the mean associated with Azzalini's centred parametrization for a skew-normal distribution on the line. Henceforth we refer to $(\mu_C, \rho, \bar{\beta}_2)$ as the circular parametrization of the WSNC distribution. Apart from the fact that in the modelling of circular data the circular parameters may well be of interest in their own right, a major advantage of this parametrization is that the moment estimates of μ_C, ρ and $\bar{\beta}_2$, i.e. $\bar{\theta}$, $\bar{R}$ and $\bar{b}_2$, always provide estimates which lie within the admissible ranges specified for their population counterparts.

Using (6.3.3) and (6.3.4) together with the identity

$$\bar{\beta}_2 \rho^2 = (\alpha_1^2 - \beta_1^2)\beta_2 - 2\alpha_1 \beta_1 \alpha_2, \tag{6.4.1}$$

it follows that the circular parameters can be expressed in terms of the direct parameters as

$$\mu_C = \tan^{-1}[\{\sin\xi + \Im(\delta\eta)\cos\xi\}/\{\cos\xi - \Im(\delta\eta)\sin\xi\}], \tag{6.4.2}$$

$$\rho = \omega\{1 + \Im^2(\delta\eta)\}^{1/2}, \tag{6.4.3}$$

$$\bar{\beta}_2 = \omega^4[\Im(2\delta\eta)\{1 - \Im^2(\delta\eta)\} - 2\Im(\delta\eta)]/\{1 + \Im^2(\delta\eta)\}, \tag{6.4.4}$$

where, as previously, $\delta = \lambda/(1+\lambda^2)^{1/2}$ and $\omega = e^{-\frac{1}{2}\eta^2}$.

If estimation is to be based on the initial estimation of the circular parameters but the direct parameters are of principal interest, a means of calculating the direct parameters from the circular ones is obviously required. The values of the direct parameters can be obtained from the circular ones, firstly by solving numerically for $\delta\eta$ in

$$\frac{\bar{\beta}_2}{\rho^4} = \frac{\Im(2\delta\eta)\{1 - \Im^2(\delta\eta)\} - 2\Im(\delta\eta)}{\{1 + \Im^2(\delta\eta)\}^3}. \tag{6.4.5}$$

The value of η can then be obtained using

$$\eta = [-2\log_e \rho + \log_e\{1 + \Im^2(\delta\eta)\}]^{1/2}, \tag{6.4.6}$$

and thence that of δ. The value of $\xi \pmod{2\pi}$ is that solution to

$$\tan(\xi) = \{\tan\mu_C - \Im(\delta\eta)\}/\{1 + \Im(\delta\eta)\tan\mu_C\},$$

which satisfies (6.3.3) or, equivalently, (6.3.4). Given the role of the location parameter ξ, one might expect the performance of this sequential approach to estimating the direct parameters to be reasonable for close to symmetric parent populations. However, as λ tends to $\pm\infty$ we might equally expect its performance to deteriorate.

Results for the moment based estimation of the mean direction, mean resultant length and second central sine moment were presented in Chapter 3. For the WSNC distribution we have (6.4.3) and (6.4.4). Using these results and those for the trigonometric moments of the WSNC distribution given in (6.3.3) and (6.3.4), together with the identities in (3.3.1) of Chapter 3, we obtain

$$\begin{aligned}
\bar{\alpha}_2 &= \omega^4\{1 - \Im^2(\delta\eta) + 2\Im(\delta\eta)\Im(2\delta\eta)\}/\{1 + \Im^2(\delta\eta)\}, \\
\bar{\alpha}_3 &= \omega^9[1 - 3\Im^2(\delta\eta) + \Im(\delta\eta)\Im(3\delta\eta)\{3 - \Im^2(\delta\eta)\}]/\{1 + \Im^2(\delta\eta)\}^{3/2}, \\
\bar{\beta}_3 &= \omega^9[\Im(\delta\eta)\{\Im^2(\delta\eta) - 3\} + \Im(3\delta\eta)\{1 - 3\Im^2(\delta\eta)\}]/\{1 + \Im^2(\delta\eta)\}^{3/2}, \\
\bar{\alpha}_4 &= \frac{\omega^{16}[1 + \Im^2(\delta\eta)\{\Im^2(\delta\eta) - 6\} + 4\Im(\delta\eta)\Im(4\delta\eta)\{1 - \Im^2(\delta\eta)\}]}{\{1 + \Im^2(\delta\eta)\}^2}.
\end{aligned} \tag{6.4.7}$$

Thus, in order to apply the large-sample inferential procedures of Chapter 3 for an underlying distribution assumed to be WSNC, one first requires estimates for η and δ. These can be obtained using the procedure for transforming from the circular parameters to the direct ones based on (6.4.5) and (6.4.6), with $\bar{\beta}_2$ and ρ replaced by $\bar{b}_2$ and $\bar{R}$, respectively.

6.4.3 Sampling Properties of the Circular Parameter Estimates

In order to investigate the adequacy of the theoretical results of Chapter 3 when applied to the WSNC distribution, we designed a Monte Carlo experiment in which simulation was conducted for a wide range of (n, ρ, λ) combinations. For each such combination considered we simulated large numbers of samples of size n from a WSNC distribution with skewness parameter λ and mean resultant length ρ. In the following three subsections we summarize the findings obtained from this simulation study regarding: i) the shapes of the sampling distributions of the

estimates, ii) the validity of the theoretical asymptotic results for the bias and variance of the estimates, and iii) the efficacy of bias-correction.

6.4.3.1 Sampling Distributions of the Individual Parameter Estimates

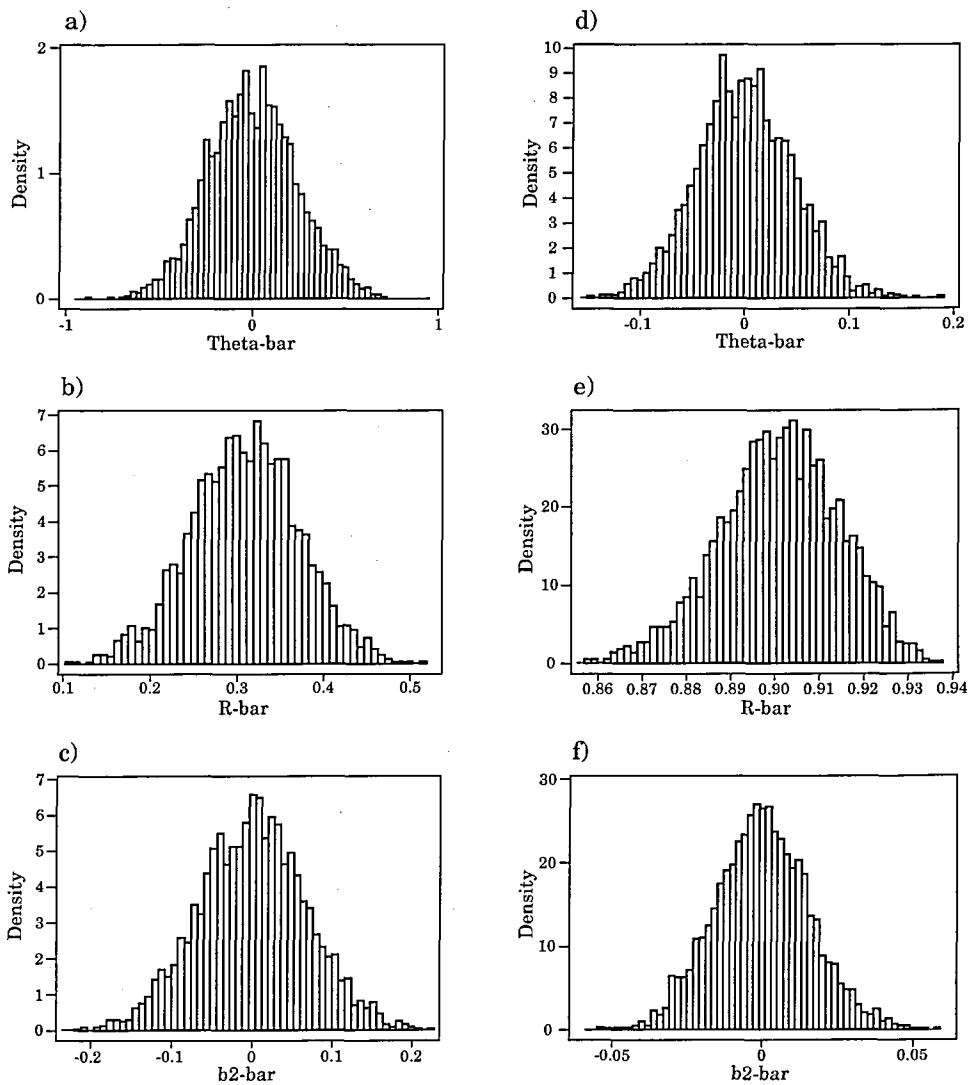

Figure 6.3 Empirical sampling distributions of the method of moments estimates $\overline{\theta}$, $\overline{R}$ and $\overline{b}_2$ obtained from 3000 simulated samples of size 100 from the WSNC distribution with $\xi = 0$, $\lambda = 0$ and: a), b), c) $\rho = 0.3$; d), e), f) $\rho = 0.9$.

In Figures 6.3 and 6.4 we illustrate the forms assumed by the sampling distributions of $\overline{\theta}$, $\overline{R}$ and $\overline{b}_2$. Each set of three histograms in these figures is based on the estimates obtained from 3000 samples of size n = 100 simulated from a WSNC distributions with $\xi = 0$ and, for those on the left-hand side of each figure, $\rho = 0.3$, and, for those on the right-hand side, $\rho = 0.9$. The samples used to produce Figure 6.3 were simulated from the symmetric wrapped normal distribution, whilst

those on which Figure 6.4 is based were simulated from the highly skew case of the WSNC distribution corresponding to a value of the skewness parameter of $\lambda = 20$. As can be appreciated from Figure 6.2, this latter distribution is very similar in form to a wrapped half-normal distribution. Even for such a highly skew parent population, as we can see from Figure 6.4, for a sample size of 100 the normal distribution provides a reasonable approximation to the sampling distributions of all three estimators. The most skew of the sampling distributions is that of $\overline{b}_2$ corresponding to the highly skew and highly concentrated case of the distribution with $\lambda = 20$ and $\rho = 0.9$.

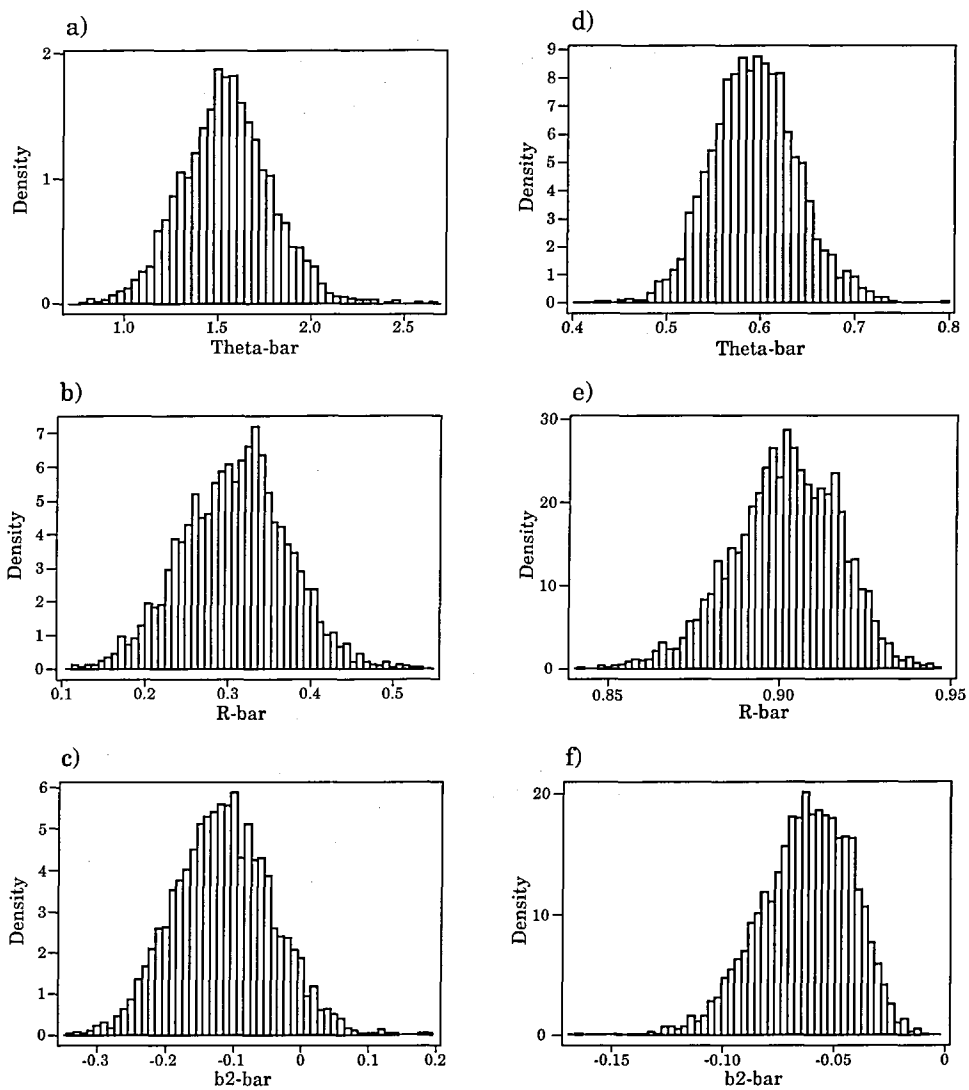

Figure 6.4 Empirical sampling distributions of the method of moments estimates $\overline{\theta}$, $\overline{R}$ and $\overline{b}_2$ obtained from 3000 simulated samples of size 100 from the WSNC distribution with $\xi = 0$, $\lambda = 20$ and: a), b), c) $\rho = 0.3$; d), e), f) $\rho = 0.9$.

In Figures 6.5 and 6.6 we present sampling distributions analogous to those presented in Figures 6.3 and 6.4, obtained using simulated samples of size 20 instead of 100. Although for such small sized samples the sampling distributions are not generally so well approximated by normal distributions, the sampling distributions of all three estimates are still, nevertheless, unimodal. However, we note that, as ρ tends further towards 0, the sampling distribution of $\overline{\theta}$ tends to the uniform distribution on $(0, 2\pi)$. Comparing the sampling distributions corresponding to ρ = 0.3 and ρ = 0.9. in both figures, we see that the degree of skewness of the sampling distributions of $\overline{R}$ and $\overline{b}_2$ depends heavily on the concentration of the parent population.

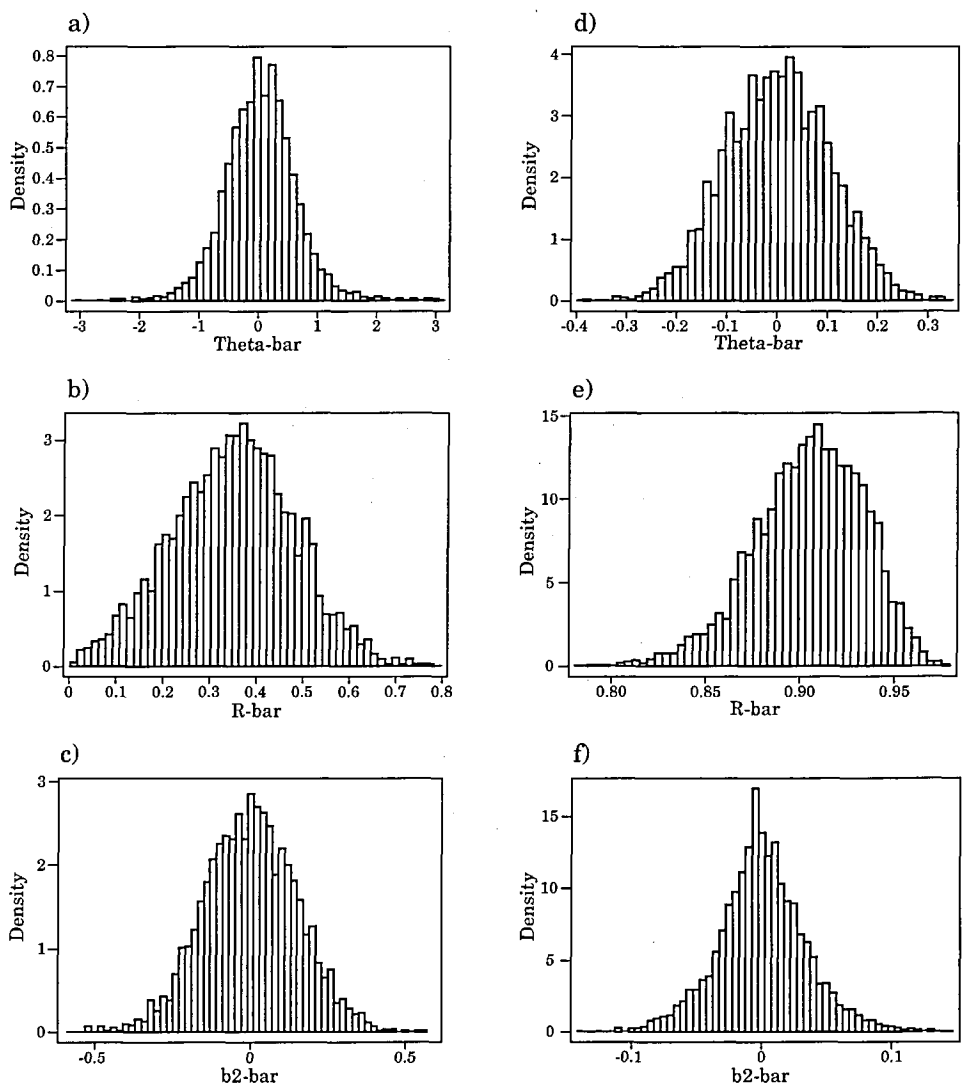

Figure 6.5 Empirical sampling distributions of the method of moments estimates $\overline{\theta}$, $\overline{R}$ and $\overline{b}_2$ obtained from 3000 simulated samples of size 20 from the WSNC distribution with $\xi = 0$, $\lambda = 0$ and: a), b), c) $\rho = 0.3$; d), e), f) $\rho = 0.9$.

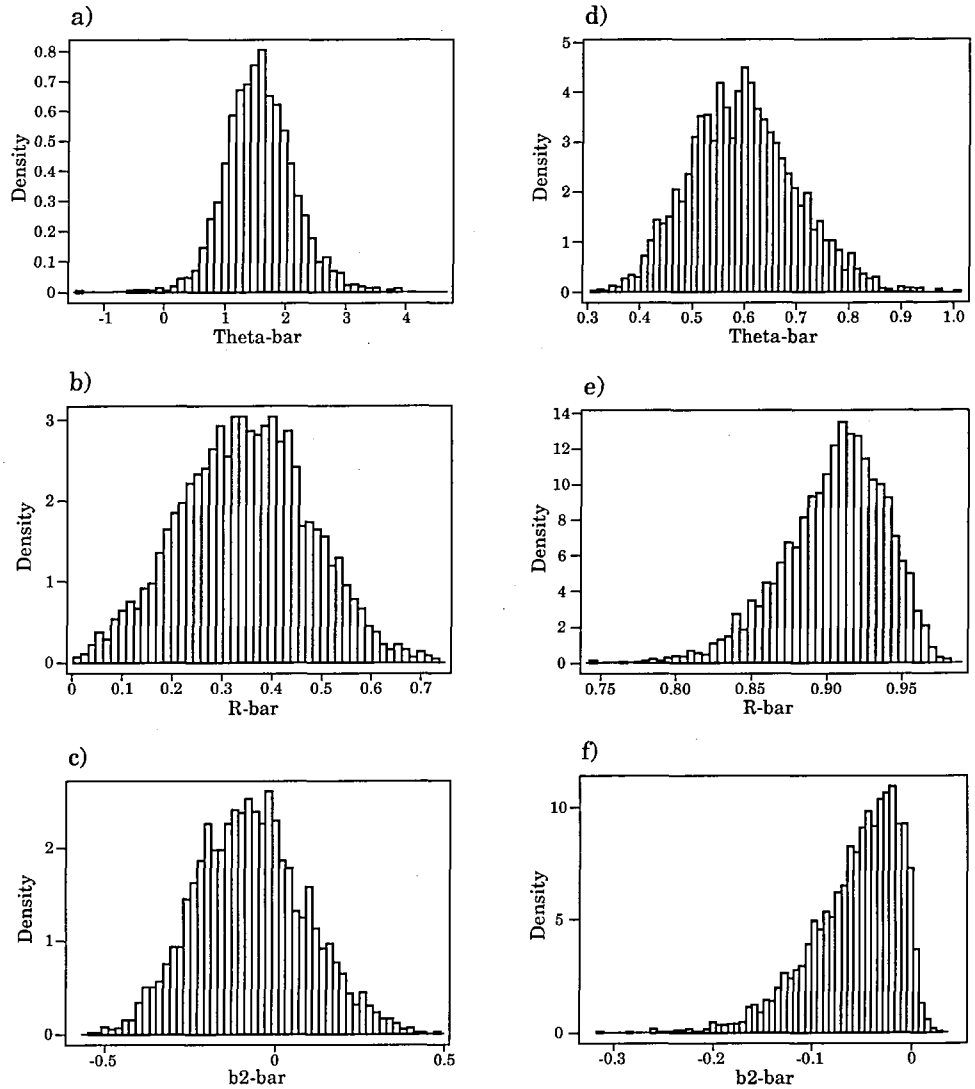

Figure 6.6 Empirical sampling distributions of the method of moments estimates $\overline{\theta}$, $\overline{R}$ and $\overline{b}_2$ obtained from 3000 simulated samples of size 20 from the WSNC distribution with $\xi = 0$, $\lambda = 20$ and: a), b), c) $\rho = 0.3$; d), e), f) $\rho = 0.9$.

6.4.3.2 Validity of the Theoretical Asymptotic Bias and Variance Results

In Tables 6.2-6.7 we present empirical results for the biases and variances of $\overline{\theta}$, $\overline{R}$ and $\overline{b}_2$. Also given are the equivalent theoretical asymptotic values calculated using (3.3.1) and (3.3.2) of Chapter 3 in conjunction with the central trigonometric moments for the WSNC distribution given in (6.4.7).

Table 6.2 Empirical and, in brackets, theoretical asymptotic bias of $\bar{\theta}$ quoted to four decimal places. Each empirical bias estimate is based on 5000 simulated samples of size n from the WSNC distribution with mean direction 0, mean resultant length ρ and skewness parameter λ.

		ρ				
λ	n	0.1	0.3	0.5	0.7	0.9
0	20	−0.0270(0.0000)	0.0046(0.0000)	0.0083(0.0000)	−0.0029(0.0000)	0.0007(0.0000)
	30	0.0095(0.0000)	0.0049(0.0000)	−0.0008(0.0000)	0.0003(0.0000)	−0.0001(0.0000)
	50	0.0052(0.0000)	−0.0010(0.0000)	−0.0006(0.0000)	−0.0009(0.0000)	−0.0004(0.0000)
	100	−0.0089(0.0000)	0.0067(0.0000)	0.0010(0.0000)	0.0014(0.0000)	0.0001(0.0000)
	200	−0.0100(0.0000)	0.0012(0.0000)	0.0025(0.0000)	−0.0010(0.0000)	0.0003(0.0000)
	500	0.0022(0.0000)	0.0013(0.0000)	0.0001(0.0000)	−0.0002(0.0000)	−0.0002(0.0000)
2	20	0.0040(0.0036)	0.0068(0.0061)	0.0114(0.0056)	0.0004(0.0036)	0.0026(0.0009)
	30	0.0049(0.0024)	0.0052(0.0041)	−0.0055(0.0038)	−0.0027(0.0024)	−0.0006(0.0006)
	50	0.0025(0.0014)	−0.0001(0.0025)	0.0037(0.0023)	0.0018(0.0014)	0.0011(0.0004)
	100	−0.0034(0.0007)	−0.0007(0.0012)	0.0015(0.0011)	−0.0015(0.0007)	−0.0005(0.0002)
	200	−0.0035(0.0004)	−0.0009(0.0006)	0.0004(0.0006)	0.0005(0.0004)	0.0004(0.0001)
	500	−0.0025(0.0001)	−0.0004(0.0002)	0.0004(0.0002)	−0.0003(0.0001)	−0.0002(0.0000)
5	20	−0.0194(0.0265)	0.0130(0.0222)	0.0085(0.0139)	0.0046(0.0074)	0.0014(0.0017)
	30	0.0106(0.0176)	0.0154(0.0148)	0.0108(0.0093)	0.0057(0.0049)	0.0011(0.0011)
	50	0.0218(0.0106)	0.0048(0.0089)	0.0059(0.0056)	0.0047(0.0030)	0.0007(0.0007)
	100	−0.0047(0.0053)	0.0012(0.0044)	0.0003(0.0028)	0.0022(0.0015)	0.0000(0.0003)
	200	−0.0017(0.0026)	0.0080(0.0022)	−0.0001(0.0014)	−0.0004(0.0007)	0.0000(0.0002)
	500	−0.0012(0.0011)	−0.0013(0.0009)	0.0012(0.0006)	0.0002(0.0003)	−0.0004(0.0000)
20	20	−0.0244(0.0995)	0.0376(0.0353)	0.0211(0.0180)	0.0098(0.0089)	0.0015(0.0020)
	30	−0.0042(0.0663)	0.0150(0.0235)	0.0118(0.0120)	0.0066(0.0059)	0.0027(0.0013)
	50	0.0011(0.0398)	0.0131(0.0141)	0.0117(0.0072)	0.0042(0.0036)	0.0010(0.0008)
	100	−0.0012(0.0199)	0.0045(0.0071)	0.0030(0.0036)	0.0020(0.0018)	0.0002(0.0004)
	200	0.0027(0.0099)	0.0068(0.0035)	0.0017(0.0018)	0.0017(0.0009)	0.0002(0.0002)
	500	0.0051(0.0040)	0.0010(0.0014)	−0.0010(0.0007)	0.0000(0.0004)	−0.0002(0.0001)

From a consideration of the results presented in these six tables, it can be appreciated that the agreement between the empirical and theoretical results is generally excellent. The largest differences between the two types of result are those for the variance of $\bar{\theta}$ and, for an underlying distribution which is highly skew, the bias of $\bar{b}_2$. As is to be expected, the disparities are largest for small sized samples drawn from highly dispersed cases of the distribution. The differences between the empirical and theoretical results under these conditions should be no surprise to us, firstly because as $\rho \to 0$ the WSNC distribution tends to the uniform distribution, the mean direction of which is undefined, and, secondly, because the theoretical results under consideration are asymptotic ones.

Table 6.3 Empirical and, in brackets, theoretical asymptotic variance of $\bar{\theta}$ quoted to four decimal places. Each empirical variance estimate is based on 5000 simulated samples of size n from the WSNC distribution with mean direction 0, mean resultant length ρ and skewness parameter λ.

		ρ				
λ	n	0.1	0.3	0.5	0.7	0.9
0	20	1.8420(2.4998)	0.3807(0.2755)	0.0967(0.0938)	0.0384(0.0388)	0.0104(0.0106)
	30	1.5790(1.6665)	0.2398(0.1837)	0.0652(0.0625)	0.0256(0.0258)	0.0070(0.0071)
	50	1.2869(0.9999)	0.1233(0.1102)	0.0368(0.0375)	0.0152(0.0155)	0.0042(0.0042)
	100	0.8183(0.5000)	0.0577(0.0551)	0.0192(0.0188)	0.0078(0.0078)	0.0022(0.0021)
	200	0.3691(0.2500)	0.0280(0.0280)	0.0093(0.0094)	0.0038(0.0039)	0.0010(0.0011)
	500	0.1136(0.1000)	0.0112(0.0110)	0.0039(0.0038)	0.0015(0.0016)	0.0004(0.0004)
2	20	1.8511(2.4998)	0.3634(0.2747)	0.0980(0.0925)	0.0387(0.0380)	0.0105(0.0105)
	30	1.6189(1.6665)	0.2375(0.1831)	0.0628(0.0617)	0.0258(0.0253)	0.0071(0.0070)
	50	1.2106(0.9999)	0.1231(0.1099)	0.0388(0.0370)	0.0151(0.0152)	0.0040(0.0042)
	100	0.7615(0.5000)	0.0567(0.0549)	0.0186(0.0185)	0.0078(0.0076)	0.0022(0.0021)
	200	0.3574(0.2500)	0.0277(0.0275)	0.0091(0.0093)	0.0038(0.0038)	0.0010(0.0010)
	500	0.1136(0.1000)	0.0107(0.0110)	0.0037(0.0037)	0.0015(0.0015)	0.0004(0.0004)
5	20	1.7854(2.5000)	0.3783(0.2750)	0.0997(0.0915)	0.0368(0.0371)	0.0102(0.0103)
	30	1.6012(1.6667)	0.2269(0.1834)	0.0652(0.0610)	0.0253(0.0247)	0.0068(0.0069)
	50	1.2351(1.0000)	0.1222(0.1100)	0.0367(0.0366)	0.0154(0.0148)	0.0040(0.0041)
	100	0.7482(0.5000)	0.0563(0.0550)	0.0176(0.0183)	0.0077(0.0074)	0.0021(0.0021)
	200	0.3660(0.2500)	0.0278(0.0275)	0.0093(0.0092)	0.0037(0.0037)	0.0010(0.0010)
	500	0.1147(0.1000)	0.0111(0.0110)	0.0037(0.0037)	0.0015(0.0015)	0.0004(0.0004)
20	20	1.7955(2.5000)	0.3625(0.2757)	0.0985(0.0913)	0.0360(0.0368)	0.0104(0.0103)
	30	1.5767(1.6667)	0.2332(0.1838)	0.0631(0.0609)	0.0254(0.0245)	0.0069(0.0068)
	50	1.3161(1.0000)	0.1195(0.1103)	0.0379(0.0365)	0.0149(0.0147)	0.0043(0.0041)
	100	0.7224(0.5000)	0.0592(0.0551)	0.0181(0.0183)	0.0073(0.0074)	0.0020(0.0021)
	200	0.3652(0.2500)	0.0280(0.0276)	0.0093(0.0091)	0.0037(0.0037)	0.0010(0.0010)
	500	0.1168(0.1000)	0.0110(0.0110)	0.0035(0.0037)	0.0015(0.0015)	0.0004(0.0004)

6.4.3.3 The Use of Bias-correction

In Section 3.4 of Chapter 3 we proposed general expressions for bias-corrected estimators which are relevant to our current deliberations. Although it would be possible to derive other bias-corrected estimators for μ_C, ρ and $\bar{\beta}_2$ designed specifically for use with an assumed underlying WSNC distribution, we decided to use simulation to compare the performance of the non-bias-corrected method of moments estimators $\bar{\theta}$, $\bar{R}$ and $\bar{b}_2$ with that of the relevant bias-corrected estimators as originally specified in Equation (3.4.1) of Section 3.4.

Table 6.4 Empirical and, in brackets, theoretical asymptotic bias of $\overline{R}$ quoted to four decimal places. Each empirical bias estimate is based on 5000 simulated samples of size n from the WSNC distribution with mean direction 0, mean resultant length ρ and skewness parameter λ.

		ρ				
λ	n	0.1	0.3	0.5	0.7	0.9
0	20	0.1197(0.1250)	0.0423(0.0413)	0.0174(0.0234)	0.0138(0.0136)	0.0049(0.0048)
	30	0.0844(0.0833)	0.0300(0.0276)	0.0145(0.0156)	0.0091(0.0090)	0.0033(0.0032)
	50	0.0559(0.0500)	0.0178(0.0165)	0.0101(0.0094)	0.0058(0.0054)	0.0018(0.0019)
	100	0.0280(0.0250)	0.0089(0.0083)	0.0047(0.0047)	0.0030(0.0027)	0.0009(0.0010)
	200	0.0133(0.0125)	0.0050(0.0041)	0.0029(0.0023)	0.0015(0.0014)	0.0004(0.0005)
	500	0.0049(0.0050)	0.0019(0.0017)	0.0009(0.0009)	0.0006(0.0005)	0.0002(0.0002)
2	20	0.1184(0.1250)	0.0469(0.0412)	0.0256(0.0231)	0.0126(0.0133)	0.0049(0.0047)
	30	0.0851(0.0833)	0.0286(0.0275)	0.0154(0.0154)	0.0083(0.0089)	0.0035(0.0031)
	50	0.0560(0.0500)	0.0162(0.0165)	0.0094(0.0093)	0.0053(0.0053)	0.0015(0.0019)
	100	0.0273(0.0250)	0.0088(0.0082)	0.0046(0.0046)	0.0023(0.0027)	0.0008(0.0009)
	200	0.0137(0.0125)	0.0041(0.0041)	0.0020(0.0023)	0.0020(0.0013)	0.0003(0.0005)
	500	0.0049(0.0050)	0.0014(0.0016)	0.0010(0.0009)	0.0006(0.0005)	0.0002(0.0002)
5	20	0.1173(0.1250)	0.0457(0.0413)	0.0254(0.0229)	0.0142(0.0130)	0.0049(0.0046)
	30	0.0872(0.0833)	0.0319(0.0275)	0.0136(0.0153)	0.0070(0.0087)	0.0025(0.0031)
	50	0.0570(0.0500)	0.0185(0.0165)	0.0102(0.0092)	0.0040(0.0052)	0.0017(0.0019)
	100	0.0298(0.0250)	0.0079(0.0083)	0.0040(0.0046)	0.0023(0.0026)	0.0009(0.0009)
	200	0.0139(0.0125)	0.0038(0.0041)	0.0021(0.0023)	0.0015(0.0013)	0.0004(0.0005)
	500	0.0048(0.0050)	0.0023(0.0017)	0.0006(0.0009)	0.0005(0.0005)	0.0003(0.0002)
20	20	0.1200(0.1250)	0.0429(0.0414)	0.0211(0.0228)	0.0122(0.0129)	0.0050(0.0046)
	30	0.0845(0.0833)	0.0256(0.0276)	0.0143(0.0152)	0.0075(0.0086)	0.0026(0.0031)
	50	0.0537(0.0500)	0.0177(0.0165)	0.0061(0.0091)	0.0045(0.0052)	0.0015(0.0018)
	100	0.0291(0.0250)	0.0099(0.0083)	0.0044(0.0046)	0.0034(0.0026)	0.0008(0.0009)
	200	0.0132(0.0125)	0.0045(0.0041)	0.0026(0.0023)	0.0003(0.0013)	0.0004(0.0005)
	500	0.0062(0.0050)	0.0014(0.0017)	0.0010(0.0009)	0.0008(0.0005)	0.0002(0.0002)

For each of a wide range of (n, ρ, λ) combinations, we simulated 3000 samples of size n from the WSNC distribution with skewness parameter λ, mean resultant length ρ, and mean direction $\mu_C = 0$. Our comparison in performance was based on the empirical bias and MSE of the individual bias-corrected and non-bias-corrected estimates obtained for the simulated samples. Also, in an attempt to explore the types of phenomena referred to in Section 3.6 of Chapter 3, we recorded the percentage of the bias-corrected estimates of ρ and $\overline{\beta}_2$ which lay outside the admissible ranges of $[0, 1]$ and $[-1, 1]$, respectively. Rather than present the detailed results obtained, instead we summarize our general findings.

Overall, the results for bias-correction were disappointing. Although it was found generally to reduce the bias in the estimation of ρ, it did not lead to a

universal reduction in bias for the estimation of the other two parameters. More damningly, it always led to an increase in mean squared error. Consequently, the use of the bias-corrected estimators for μ_C, ρ and $\overline{\beta}_2$ cannot be recommended when the parent population is WSNC.

Table 6.5 Empirical and, in brackets, theoretical asymptotic variance of $\overline{R}$ quoted to four decimal places. Each empirical variance estimate is based on 5000 simulated samples of size n from the WSNC distribution with mean direction 0, mean resultant length ρ and skewness parameter λ.

		ρ				
λ	n	0.1	0.3	0.5	0.7	0.9
0	20	0.0122(0.0245)	0.0167(0.0207)	0.0130(0.0141)	0.0062(0.0065)	0.0009(0.0009)
	30	0.0088(0.0163)	0.0119(0.0138)	0.0089(0.0094)	0.0042(0.0043)	0.0006(0.0006)
	50	0.0060(0.0098)	0.0076(0.0083)	0.0054(0.0056)	0.0026(0.0026)	0.0004(0.0004)
	100	0.0035(0.0049)	0.0040(0.0041)	0.0028(0.0028)	0.0013(0.0013)	0.0002(0.0002)
	200	0.0020(0.0025)	0.0020(0.0021)	0.0014(0.0014)	0.0006(0.0007)	0.0001(0.0001)
	500	0.0010(0.0010)	0.0008(0.0008)	0.0005(0.0006)	0.0003(0.0003)	0.0000(0.0000)
2	20	0.0123(0.0245)	0.0174(0.0208)	0.0135(0.0144)	0.0069(0.0069)	0.0010(0.0010)
	30	0.0087(0.0163)	0.0122(0.0139)	0.0089(0.0096)	0.0046(0.0046)	0.0007(0.0007)
	50	0.0060(0.0098)	0.0078(0.0083)	0.0058(0.0057)	0.0027(0.0028)	0.0004(0.0004)
	100	0.0035(0.0049)	0.0041(0.0042)	0.0029(0.0029)	0.0014(0.0014)	0.0002(0.0002)
	200	0.0021(0.0025)	0.0020(0.0021)	0.0015(0.0014)	0.0007(0.0007)	0.0001(0.0001)
	500	0.0009(0.0010)	0.0009(0.0008)	0.0006(0.0006)	0.0003(0.0003)	0.0000(0.0000)
5	20	0.0117(0.0245)	0.0174(0.0207)	0.0142(0.0146)	0.0069(0.0073)	0.0010(0.0011)
	30	0.0091(0.0163)	0.0120(0.0138)	0.0098(0.0097)	0.0047(0.0049)	0.0007(0.0008)
	50	0.0059(0.0098)	0.0076(0.0083)	0.0057(0.0058)	0.0030(0.0029)	0.0004(0.0005)
	100	0.0036(0.0049)	0.0042(0.0041)	0.0030(0.0029)	0.0015(0.0015)	0.0002(0.0002)
	200	0.0020(0.0025)	0.0020(0.0021)	0.0015(0.0015)	0.0007(0.0007)	0.0001(0.0001)
	500	0.0009(0.0010)	0.0009(0.0008)	0.0006(0.0006)	0.0003(0.0003)	0.0000(0.0000)
20	20	0.0121(0.0245)	0.0172(0.0207)	0.0141(0.0147)	0.0073(0.0075)	0.0011(0.0012)
	30	0.0086(0.0163)	0.0122(0.0138)	0.0093(0.0098)	0.0048(0.0050)	0.0008(0.0008)
	50	0.0058(0.0098)	0.0076(0.0083)	0.0057(0.0059)	0.0029(0.0030)	0.0005(0.0005)
	100	0.0035(0.0049)	0.0040(0.0041)	0.0028(0.0029)	0.0015(0.0015)	0.0002(0.0002)
	200	0.0021(0.0025)	0.0020(0.0021)	0.0014(0.0015)	0.0007(0.0007)	0.0001(0.0001)
	500	0.0009(0.0010)	0.0008(0.0008)	0.0006(0.0006)	0.0003(0.0003)	0.0000(0.0000)

We conclude our observations regarding bias-correction by noting that the frequencies of inadmissible bias-corrected estimates for ρ and $\overline{\beta}_2$ were greatest for small-sized samples drawn from highly dispersed cases of the WSNC distribution. The frequencies of inadmissible bias-corrected estimates for $\overline{\beta}_2$ were consistently higher than their counterparts for ρ. Whilst the latter frequencies decreased rapidly with increasing sample size and concentration of the parent population, the frequencies of inadmissible bias-corrected estimates of $\overline{\beta}_2$ decreased much more

slowly, with an increase in concentration of the parent population leading to a more pronounced decrease in such frequencies than an increase in sample size. The skewness of the parent population was found to have only a marginal effect on both sets of frequencies.

Table 6.6 Empirical and, in brackets, theoretical asymptotic bias of $\overline{b}_2$ quoted to four decimal places. Each empirical bias estimate is based on 5000 simulated samples of size n from the WSNC distribution with mean direction 0, mean resultant length ρ and skewness parameter λ.

		ρ				
λ	n	0.1	0.3	0.5	0.7	0.9
0	20	−0.0022(0.0000)	0.0013(0.0000)	−0.0011(0.0000)	−0.0008(0.0000)	−0.0001(0.0000)
	30	0.0008(0.0000)	−0.0014(0.0000)	0.0004(0.0000)	0.0015(0.0000)	−0.0005(0.0000)
	50	−0.0004(0.0000)	0.0001(0.0000)	−0.0002(0.0000)	0.0016(0.0000)	0.0002(0.0000)
	100	−0.0004(0.0000)	0.0008(0.0000)	0.0001(0.0000)	0.0010(0.0000)	−0.0003(0.0000)
	200	0.0002(0.0000)	0.0010(0.0000)	0.0002(0.0000)	−0.0002(0.0000)	0.0000(0.0000)
	500	0.0003(0.0000)	0.0001(0.0000)	0.0007(0.0000)	0.0003(0.0000)	0.0000(0.0000)
2	20	0.0068(0.0071)	0.0164(0.0093)	0.0113(0.0058)	0.0044(0.0039)	0.0032(0.0028)
	30	0.0021(0.0047)	0.0082(0.0062)	0.0060(0.0039)	0.0028(0.0026)	0.0026(0.0019)
	50	0.0018(0.0028)	0.0072(0.0037)	0.0061(0.0023)	0.0025(0.0015)	0.0012(0.0011)
	100	0.0012(0.0014)	0.0041(0.0019)	0.0023(0.0012)	0.0010(0.0008)	0.0007(0.0006)
	200	0.0012(0.0007)	0.0020(0.0009)	0.0005(0.0006)	0.0006(0.0004)	0.0004(0.0003)
	500	0.0004(0.0003)	−0.0002(0.0004)	0.0007(0.0002)	0.0002(0.0002)	0.0000(0.0001)
5	20	0.0076(0.0529)	0.0419(0.0355)	0.0281(0.0141)	0.0147(0.0089)	0.0074(0.0055)
	30	0.0092(0.0353)	0.0295(0.0236)	0.0198(0.0094)	0.0115(0.0060)	0.0036(0.0036)
	50	0.0086(0.0212)	0.0166(0.0142)	0.0116(0.0056)	0.0070(0.0036)	0.0019(0.0022)
	100	0.0073(0.0106)	0.0079(0.0071)	0.0049(0.0028)	0.0040(0.0018)	0.0011(0.0011)
	200	0.0044(0.0053)	0.0049(0.0035)	0.0025(0.0014)	0.0018(0.0009)	0.0005(0.0005)
	500	0.0025(0.0021)	0.0016(0.0014)	0.0011(0.0006)	0.0009(0.0004)	0.0004(0.0002)
20	20	0.0393(0.1989)	0.0567(0.0601)	0.0342(0.0187)	0.0198(0.0113)	0.0079(0.0064)
	30	0.0353(0.1326)	0.0429(0.0401)	0.0247(0.0125)	0.0127(0.0075)	0.0047(0.0043)
	50	0.0322(0.0796)	0.0280(0.0240)	0.0142(0.0075)	0.0090(0.0045)	0.0028(0.0026)
	100	0.0259(0.0398)	0.0152(0.0120)	0.0081(0.0037)	0.0041(0.0023)	0.0016(0.0013)
	200	0.0182(0.0199)	0.0069(0.0060)	0.0040(0.0019)	0.0015(0.0011)	0.0007(0.0006)
	500	0.0081(0.0080)	0.0032(0.0024)	0.0016(0.0007)	0.0008(0.0005)	0.0004(0.0003)

6.4.4 Sampling Distributions of the Corresponding Direct and Centred Parameter Estimates

As mentioned in Section 6.4.2, the circular parameters of the WSNC will often be of interest in their own right. However, if, for instance, we are interested in properties related to the density of the distribution then, as the density of the WSNC cannot be expressed analytically in terms of the circular parameters, it is necessary to work with the corresponding direct or centred parameters. Here and in the remainder of the chapter we will denote the centred parameters of the general

skew-normal distribution defined on the real line as μ_L, σ and γ_1, the subindex L of μ_L being used to differentiate the parameter considered from the mean direction, μ_C, of the WSNC distribution.

Table 6.7 Empirical and, in brackets, theoretical asymptotic variance of $\bar{b}_2$ quoted to four decimal places. Each empirical variance estimate is based on 5000 simulated samples of size n from the WSNC distribution with mean direction 0, mean resultant length ρ and skewness parameter λ.

		ρ				
λ	n	0.1	0.3	0.5	0.7	0.9
0	20	0.0237(0.0250)	0.0232(0.0243)	0.0195(0.0202)	0.0106(0.0112)	0.0012(0.0013)
	30	0.0153(0.0167)	0.0153(0.0162)	0.0130(0.0135)	0.0072(0.0075)	0.0008(0.0009)
	50	0.0099(0.0100)	0.0097(0.0097)	0.0079(0.0081)	0.0043(0.0045)	0.0005(0.0005)
	100	0.0050(0.0050)	0.0047(0.0049)	0.0041(0.0040)	0.0021(0.0022)	0.0003(0.0003)
	200	0.0025(0.0025)	0.0024(0.0024)	0.0020(0.0020)	0.0011(0.0011)	0.0001(0.0001)
	500	0.0010(0.0010)	0.0010(0.0010)	0.0008(0.0008)	0.0005(0.0004)	0.0001(0.0001)
2	20	0.0237(0.0250)	0.0221(0.0240)	0.0185(0.0194)	0.0108(0.0108)	0.0014(0.0016)
	30	0.0158(0.0167)	0.0155(0.0160)	0.0125(0.0130)	0.0069(0.0072)	0.0010(0.0010)
	50	0.0099(0.0100)	0.0094(0.0096)	0.0072(0.0078)	0.0043(0.0043)	0.0006(0.0006)
	100	0.0049(0.0050)	0.0048(0.0048)	0.0038(0.0039)	0.0022(0.0022)	0.0003(0.0003)
	200	0.0025(0.0025)	0.0024(0.0024)	0.0020(0.0019)	0.0011(0.0011)	0.0002(0.0002)
	500	0.0010(0.0010)	0.0009(0.0010)	0.0008(0.0008)	0.0004(0.0004)	0.0001(0.0001)
5	20	0.0237(0.0250)	0.0236(0.0238)	0.0184(0.0181)	0.0101(0.0102)	0.0017(0.0020)
	30	0.0159(0.0167)	0.0161(0.0158)	0.0121(0.0120)	0.0067(0.0068)	0.0012(0.0013)
	50	0.0099(0.0100)	0.0096(0.0095)	0.0071(0.0072)	0.0040(0.0041)	0.0008(0.0008)
	100	0.0050(0.0050)	0.0049(0.0048)	0.0036(0.0036)	0.0020(0.0020)	0.0004(0.0004)
	200	0.0025(0.0025)	0.0024(0.0024)	0.0018(0.0018)	0.0010(0.0010)	0.0002(0.0002)
	500	0.0010(0.0010)	0.0010(0.0010)	0.0007(0.0007)	0.0004(0.0004)	0.0001(0.0001)
20	20	0.0242(0.0249)	0.0263(0.0235)	0.0187(0.0175)	0.0099(0.0101)	0.0019(0.0021)
	30	0.0166(0.0166)	0.0179(0.0156)	0.0119(0.0117)	0.0064(0.0067)	0.0014(0.0014)
	50	0.0104(0.0100)	0.0103(0.0094)	0.0070(0.0070)	0.0041(0.0040)	0.0008(0.0009)
	100	0.0053(0.0050)	0.0049(0.0047)	0.0035(0.0035)	0.0021(0.0020)	0.0004(0.0004)
	200	0.0029(0.0025)	0.0024(0.0023)	0.0018(0.0017)	0.0011(0.0010)	0.0002(0.0002)
	500	0.0011(0.0010)	0.0010(0.0009)	0.0007(0.0007)	0.0004(0.0004)	0.0001(0.0001)

If $Y_C \sim \mathrm{SN}_C(\mu_L, \sigma, \gamma_1)$ then, using (1.4.6) of Chapter 1, the circular random variable $\theta_C = Y_C \pmod{2\pi}$ has the density, expressed as a function of the centred parameters, given by

$$f(\theta;\mu_L,\sigma,\gamma_1)=\frac{2}{\sigma\left(1+c^2\gamma_1^{2/3}\right)^{1/2}}\sum_{r=-\infty}^{\infty}\phi\left\{\frac{1}{\left(1+c^2\gamma_1^{2/3}\right)^{1/2}}\left(\frac{\theta+2\pi r-\mu_L}{\sigma}+c\gamma_1^{1/3}\right)\right\}$$
$$\times\Phi\left[\frac{c\gamma_1^{1/3}}{\left\{b^2+c^2\left(b^2-1\right)\gamma_1^{2/3}\right\}^{1/2}\left(1+c^2\gamma_1^{2/3}\right)^{1/2}}\left(\frac{\theta+2\pi r-\mu_L}{\sigma}+c\gamma_1^{1/3}\right)\right], \tag{6.4.8}$$

where $0\le\theta<2\pi, -\infty<\mu_L<\infty, \sigma>0, -0.99527<\gamma_1<0.99527$, $b=(2/\pi)^{1/2}$ and $c=\{2/(4-\pi)\}^{1/3}$. We will say that θ_C is distributed according to the WSNC distribution with centred parameters (μ_L,σ,γ_1), and denote the fact using the notation $\theta_C \sim \text{WSNC}_C(\mu_L,\sigma,\gamma_1)$.

Having obtained method of moments estimates of μ_C, ρ and $\overline{\beta}_2$, one can transform them to obtain estimates of the direct parameters $\xi\,(\text{mod}\,2\pi)$, η and λ as described in Section 6.4.2. Using the results in Section 1.2.3 of Chapter 1, estimates for the centred parameters can be calculated from those for the direct parameters using the relations

$$\begin{aligned}\mu_L\,(\text{mod}\,2\pi)&=\xi\,(\text{mod}\,2\pi)+b\eta\delta,\\ \sigma&=\eta\left(1-b^2\delta^2\right)^{1/2},\\ \gamma_1&=\frac{b\delta^3\left(2b^2-1\right)}{\left(1-b^2\delta^2\right)^{3/2}}.\end{aligned} \tag{6.4.9}$$

In order to illustrate the forms taken by the sampling distributions obtained using this approach, in Figures 6.7 we provide histograms of the estimates of the direct and centred parameters obtained by transforming the estimates of the circular parameters whose sampling distributions were portrayed in the three histograms appearing in the left-hand side of Figure 6.3. Despite the perfectly regular nature of those original three histograms, the corresponding sampling distributions of the direct and centred parameters are, in general, not nearly as appealing. Specifically, those for ξ, δ and γ_1 are bimodal, while those for η and σ are highly skew. Although such sampling behaviour does not necessarily inconvenience the point estimation of the relevant parameters, it clearly has important implications for other forms of inference such as confidence interval construction and hypothesis testing, at least if classical approaches to inference are to be employed. Of course, such irregular sampling behaviour is immaterial for

computer intensive methods of inference such as the bootstrap and Monte Carlo significance tests.

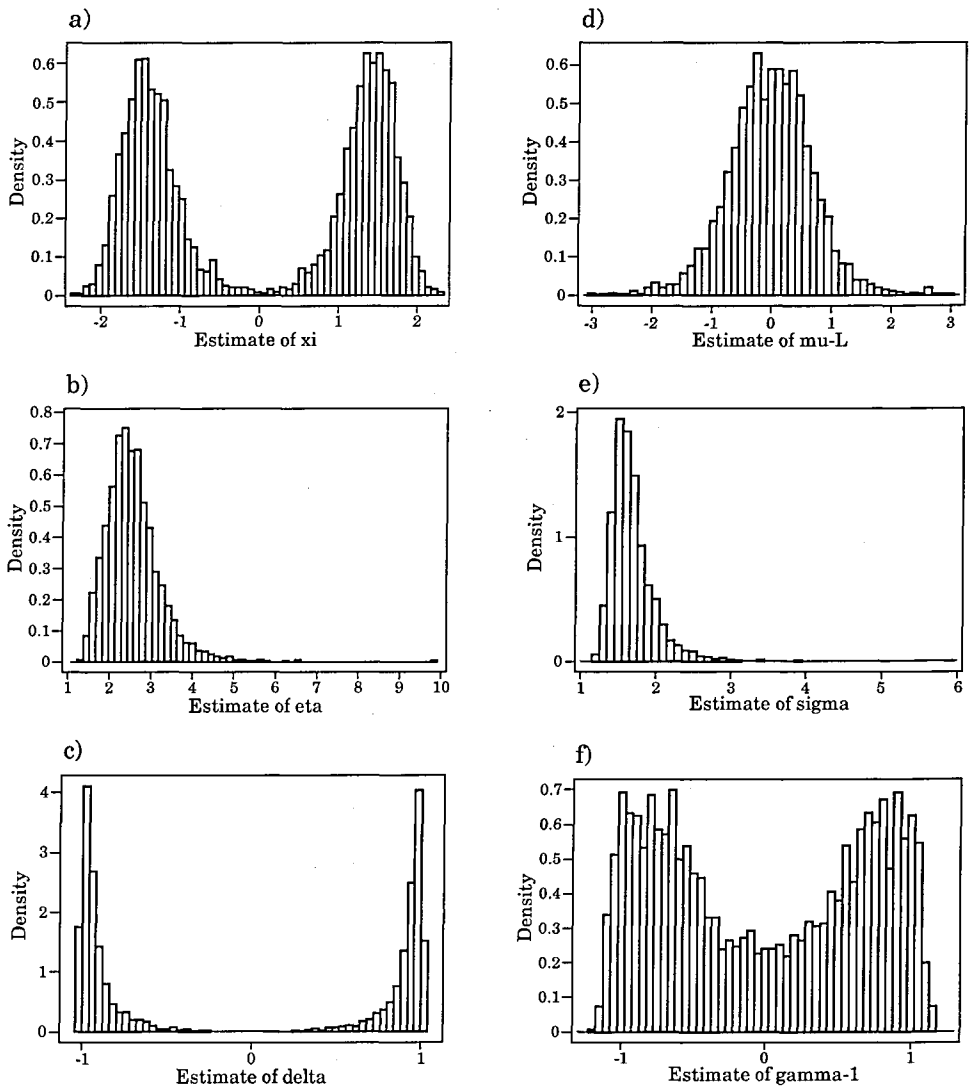

Figure 6.7 Empirical sampling distributions of the method of moments estimates of the direct parameters: a) ξ (mod 2π), b) η and c) δ, and the centred parameters: d) μ_L (mod 2π), e) σ and f) γ_1, obtained using transformation of the estimates of the circular parameters displayed in Figure 6.3a, b and c.

6.5 Maximum Likelihood Estimation

From our work on moment based inference in the previous section, there are three parametrizations which one might contemplate using within the general framework of maximum likelihood (ML) estimation. The circular parametrization rules itself out as the likelihood function cannot be expressed analytically as a function of the circular parameters. In addition, it is not generally advisable to use

the direct parametrization for the reasons related to its parameter redundancy for the normal distribution discussed in Section 1.4.1.2 of Chapter 1. Thus, as on the line, we advocate the use of the centred parametrization.

Using (6.4.8), it follows that, for a random sample of size n, $\underline{\theta}_C = \left(\theta_{C_1},...,\theta_{C_n}\right)$, from the WSNC distribution with centred parameters μ_L, σ and γ_1, the log-likelihood is given by

$$l\left(\mu_L,\sigma,\gamma_1;\underline{\theta}_C\right)=n\ln 2-n\ln\sigma-\frac{n}{2}\ln\left(1+c^2\gamma_1^{2/3}\right)$$
$$+\sum_{i=1}^{n}\ln\left(\sum_{r=-\infty}^{\infty}\phi\left\{\frac{1}{\left(1+c^2\gamma_1^{2/3}\right)^{1/2}}\left(\frac{\theta_{C_i}+2\pi r-\mu_L}{\sigma}+c\gamma_1^{1/3}\right)\right\}\right. \tag{6.5.1}$$
$$\left.\times\Phi\left[\frac{c\gamma_1^{1/3}}{\left\{b^2+c^2\left(b^2-1\right)\gamma_1^{2/3}\right\}^{1/2}\left(1+c^2\gamma_1^{2/3}\right)^{1/2}}\left(\frac{\theta_{C_i}+2\pi r-\mu_L}{\sigma}+c\gamma_1^{1/3}\right)\right]\right).$$

Numerical methods of optimization must be employed in order to maximize (6.5.1) and hence identify the maximum likelihood estimates. In practice, the infinite summation in (6.5.1) can be reduced to include just a small number of its central terms. The central three terms suffice for most applications, whilst the first seven central terms are more than adequate for data sampled from even the most dispersed of parent populations.

Our approach to maximizing (6.5.1) was similar to that described in Section 1.4.2.2 of Chapter 1 for the general skew-normal distribution on the line. We used the Nelder-Mead simplex combined with a grid of starting values. The use of a grid of initial values rather than, say, the estimates obtained from method of moments estimation as the sole starting values, is advisable as, firstly, the method of moments estimates of the centred parameters do not necessarily provide good initial values for the optimization process, and, secondly, because multiple maxima can occur on the log-likelihood surface. In Figure 6.8 we illustrate the general lack of any strong relation between the MM estimates of the centred parameters and their ML counterparts when the parent population is highly dispersed. The estimates displayed in the three scatterplots making up this figure are those obtained from the same 3000 simulated samples of size 100 used to derive the sampling distributions of the MM estimates portrayed in the three histograms making up the left-hand side of Figure 6.3, and those of Figure 6.7. We note, however, that as the concentration of the parent WSNC distribution increases, so

does the strength of the relationships between the MM and ML estimates of the centred parameters.

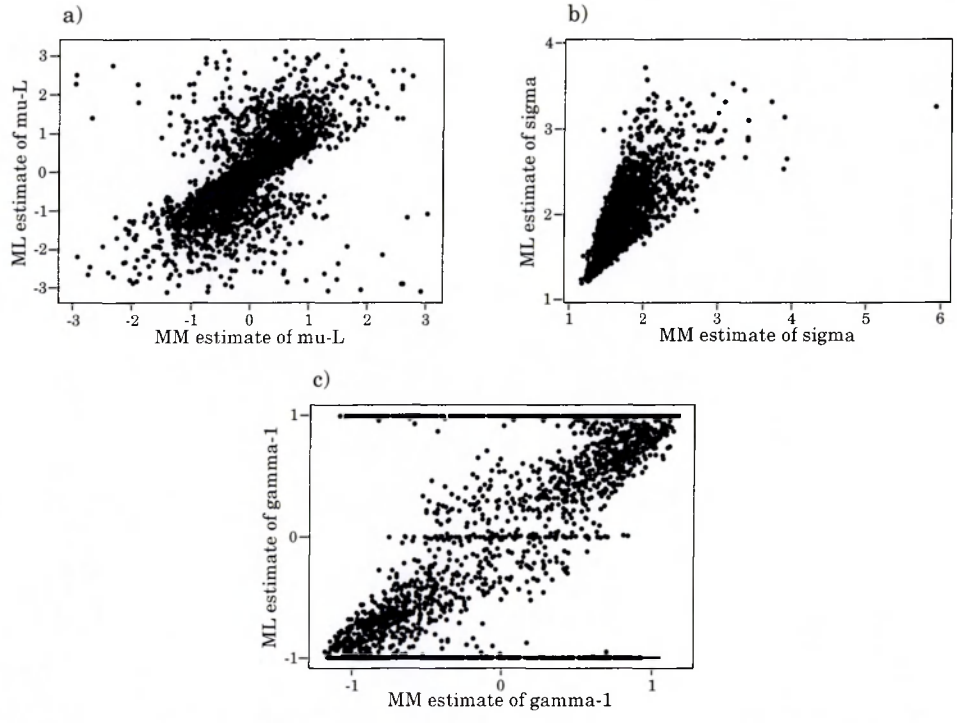

Figure 6.8 Scatterplots of the ML versus the MM estimates of the centred parameters: a) $\mu_L \pmod{2\pi}$, b) σ and c) γ_1, obtained from the same 3000 simulated samples of size 100 used in the production of the left-hand side of Figure 6.3, and Figure 6.7. The values of the centred parameters for the parent WSNC distribution are $\mu_L = 0$, $\sigma = 1.5518$ and $\gamma_1 = 0$.

During the process of optimization it is also necessary to check that the maximum of the log-likelihood function does not correspond to a wrapped half-normal distribution. For reasons of simplicity, and as parameter redundancy is not an issue on the relevant boundaries of the parameter space, this check can be carried out using numerical optimization of the log-likelihood function expressed in terms of the direct parameters with λ set equal to $\pm\infty$. This representation of the log-likelihood function is easily derived from (6.3.1). Once more, only a small number of the central terms making up the infinite summation appearing in the resulting log-likelihood need to be used in the optimization process. Again, we recommend the use of a range of starting values, in this case, for ξ and η.

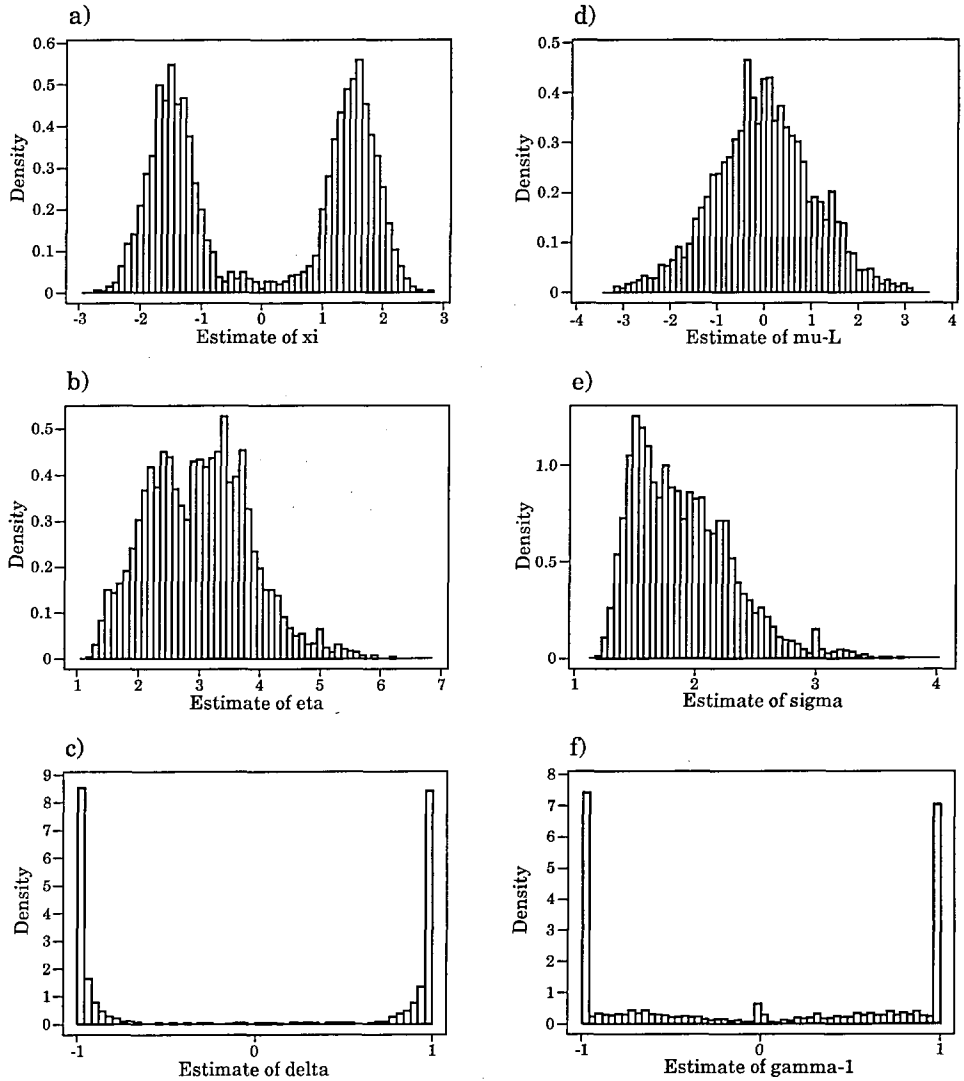

Figure 6.9 Empirical sampling distributions of the maximum likelihood estimates of the direct parameters: a) $\xi \pmod{2\pi}$, b) η and c) δ, and the centred parameters: d) $\mu_L \pmod{2\pi}$, e) σ and f) γ_1, for the same 3000 samples of size 100 used in the production of Figure 6.7.

Clearly, once ML estimation has been employed to obtain estimates of the centred parameters, these estimates can be transformed to obtain estimates of the direct or circular parameters, as and when required. The direct parameters can be calculated from the centred ones using the relations

$$\xi \left(\mathrm{mod}\, 2\pi \right) = \mu_{\mathrm{L}} \left(\mathrm{mod}\, 2\pi \right) - c\gamma_1^{1/3}\sigma,$$

$$\eta = \sigma\left(1 + c^2\gamma_1^{2/3}\right)^{1/2},$$

$$\lambda = \frac{c\gamma_1^{1/3}}{\left\{b^2 + c^2\left(b^2 - 1\right)\gamma_1^{2/3}\right\}^{1/2}}.$$

The estimates of the circular parameters can then be calculated from those for the direct parameters using Equations (6.4.2)-(6.4.4) of Section 6.4.2. We note however that the sampling distributions of the estimators for all three parametrizations are not, in general, particularly appealing. To illustrate this fact, in Figure 6.9 we display the sampling distributions of the individual ML estimates of the direct and centred parameters obtained for the same 3000 simulated samples of size 100 used previously in the production of Figures 6.3, 6.7 and 6.8. In Figure 6.10 we portray the sampling distributions of the corresponding estimates of the circular parameters. A comparison of the histograms represented in these two figures with their counterparts in Figures 6.7 and 6.3, respectively, provides an indication of the similarities, as well as the differences, that exist between the sampling distributions of the MM and ML estimates of the various parameters.

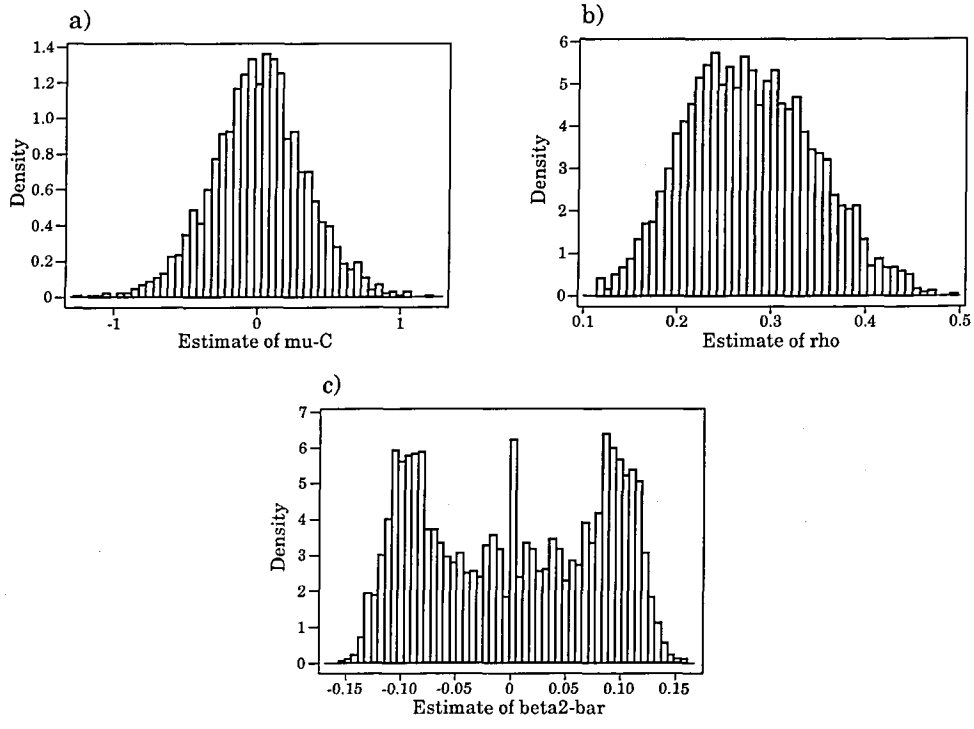

Figure 6.10 Empirical sampling distributions of the maximum likelihood estimates of the circular parameters: a) μ_C, b) ρ and c) $\overline{\beta}_2$, for the same 3000 samples of size 100 used to produce the histograms making up the left-hand side of Figure 6.3.

6.6 Detailed Empirical Comparison of MM and ML Estimation

As an extension of the Monte Carlo experiment described in Section 6.4.3, we used simulation to compare the performance of method of moments and maximum likelihood estimation in greater detail. In order to be consistent with the parametrization used in Section 1.4.2.3 of Chapter 1, in Tables 6.8-6.10 we present, individually, the results obtained for the estimates of the centred parameters. Given that this parametrization is the one advocated for use with ML estimation, whereas our approach to MM estimation is based on the circular parametrization, one might perhaps expect our decision to represent the results for the centred parameters to be favourable towards ML estimation.

Table 6.8 Bias and, in brackets, MSE for the MM and ML estimates of μ_L. For each (n, ρ, λ) combination the bias and MSE values were calculated using 3000 simulated samples of size n from the WSNC distribution with $\xi = 0$ and ρ and λ as specified.

		ρ	0.3		0.5		0.7		0.9	
λ	n		MM	ML	MM	ML	MM	ML	MM	ML
0	20		0.0308	0.0101	−0.0070	0.0091	0.0049	0.0037	0.0013	0.0001
			(1.2455)	(1.1744)	(0.1975)	(0.2437)	(0.0421)	(0.0474)	(0.0123)	(0.0112)
	50		−0.0031	0.0223	−0.0046	−0.0045	0.0057	0.0037	−0.0024	−0.0006
			(0.9240)	(1.3168)	(0.0659)	(0.1083)	(0.0170)	(0.0161)	(0.0057)	(0.0042)
	100		−0.0247	−0.0144	0.0003	0.0041	0.0017	0.0014	0.0005	0.0001
			(0.4941)	(1.1192)	(0.0328)	(0.0192)	(0.0095)	(0.0077)	(0.0036)	(0.0021)
2	20		−0.2183	−0.2494	−0.0313	−0.0473	0.0162	−0.0089	0.0198	−0.0052
			(1.3544)	(1.2750)	(0.2108)	(0.2697)	(0.0417)	(0.0490)	(0.0120)	(0.0109)
	50		−0.1793	−0.1302	−0.0124	−0.0136	0.0212	−0.0062	0.0297	−0.0033
			(0.9335)	(1.3401)	(0.0707)	(0.1216)	(0.0162)	(0.0156)	(0.0058)	(0.0043)
	100		−0.1490	−0.0941	0.0043	−0.0006	0.0265	−0.0017	0.0365	−0.0021
			(0.5334)	(1.1550)	(0.0322)	(0.0392)	(0.0091)	(0.0076)	(0.0037)	(0.0022)
5	20		−0.4773	−0.4867	−0.0342	−0.0746	0.0233	−0.0139	0.0337	0.0003
			(1.4813)	(1.3278)	(0.2463)	(0.2157)	(0.0472)	(0.0495)	(0.0122)	(0.0114)
	50		−0.2527	−0.1576	0.0004	0.0057	0.0248	−0.0082	0.0375	−0.0061
			(0.9444)	(1.1149)	(0.0733)	(0.0793)	(0.0170)	(0.0154)	(0.0055)	(0.0042)
	100		−0.1440	−0.0294	0.0099	0.0045	0.0245	−0.0039	0.0378	−0.0030
			(0.4677)	(0.6819)	(0.0329)	(0.0276)	(0.0091)	(0.0075)	(0.0035)	(0.0020)
20	20		−0.5567	−0.5185	−0.0354	−0.0403	0.0353	0.0202	0.0364	0.0093
			(1.4776)	(1.1467)	(0.2364)	(0.1312)	(0.0493)	(0.0378)	(0.0121)	(0.0105)
	50		−0.2670	−0.1585	−0.0085	−0.0165	0.0213	0.0027	0.0387	0.0005
			(0.9471)	(0.6153)	(0.0784)	(0.0428)	(0.0195)	(0.0151)	(0.0057)	(0.0040)
	100		−0.0947	−0.0338	−0.0064	−0.0001	0.0211	−0.0039	0.0375	−0.0031
			(0.4578)	(0.2536)	(0.0378)	(0.0199)	(0.0094)	(0.0070)	(0.0035)	(0.0020)

Table 6.9 Bias and, in brackets, MSE for the MM and ML estimates of σ. For each (n, ρ, λ) combination the bias and MSE values were calculated using 3000 simulated samples of size n from the WSNC distribution with $\xi = 0$ and ρ and λ as specified.

	ρ	0.3		0.5		0.7		0.9	
λ	n	MM	ML	MM	ML	MM	ML	MM	ML
0	20	0.3271	0.1015	0.0239	0.0661	0.0472	−0.0105	−0.0531	0.0032
		(1.1693)	(0.1464)	(0.1269)	(0.0730)	(0.0282)	(0.0256)	(0.0081)	(0.0062)
	50	0.2621	0.3540	0.0015	0.0689	−0.0361	−0.0008	−0.0389	−0.0021
		(0.5498)	(0.3055)	(0.0257)	(0.0448)	(0.0096)	(0.0092)	(0.0038)	(0.0042)
	100	0.1346	0.3450	0.0085	0.0241	−0.0334	−0.0063	−0.0322	−0.0027
		(0.1127)	(0.2836)	(0.0112)	(0.0162)	(0.0051)	(0.0038)	(0.0022)	(0.0011)
2	20	0.3313	0.0646	−0.0030	0.0401	−0.0511	0.0075	−0.0601	−0.0019
		(1.2978)	(0.1472)	(0.1160)	(0.0672)	(0.0278)	(0.0266)	(0.0093)	(0.0067)
	50	0.1791	0.2819	−0.0210	0.0562	−0.0393	0.0021	−0.0469	−0.0021
		(0.4463)	(0.2574)	(0.0262)	(0.0427)	(0.0111)	(0.0109)	(0.0045)	(0.0027)
	100	0.0810	0.2948	−0.0211	0.0165	−0.0412	−0.0070	−0.0438	−0.0022
		(0.1260)	(0.2593)	(0.0133)	(0.0180)	(0.0064)	(0.0044)	(0.0031)	(0.0013)
5	20	0.2261	0.1018	−0.0248	−0.0243	−0.0617	−0.0291	−0.0644	−0.0111
		(1.4629)	(0.1440)	(0.1371)	(0.0600)	(0.0336)	(0.0259)	(0.0105)	(0.0066)
	50	0.0953	0.1294	−0.0320	0.0158	−0.0457	−0.0009	−0.0573	−0.0009
		(0.5013)	(0.1971)	(0.0355)	(0.0337)	(0.0135)	(0.0108)	(0.0058)	(0.0029)
	100	−0.0100	0.1177	−0.0309	0.0071	−0.0398	−0.0016	−0.0564	−0.0016
		(0.1505)	(0.1449)	(0.0175)	(0.0172)	(0.0078)	(0.0054)	(0.0045)	(0.0015)
20	20	0.0366	−0.2473	−0.0337	−0.0846	−0.0517	−0.0490	−0.0694	−0.0288
		(1.1878)	(0.1816)	(0.1468)	(0.0548)	(0.0341)	(0.0223)	(0.0118)	(0.0065)
	50	−0.0765	−0.0384	−0.0447	−0.0306	−0.0472	−0.0146	−0.0629	−0.0080
		(0.6889)	(0.1250)	(0.0784)	(0.0223)	(0.0151)	(0.0092)	(0.0067)	(0.0024)
	100	−0.0135	−0.0050	−0.0380	−0.0066	−0.0381	−0.0024	−0.0609	−0.0027
		(0.1942)	(0.0739)	(0.0207)	(0.0119)	(0.0094)	(0.0044)	(0.0051)	(0.0013)

The results in the three tables are heavily conditioned by the fact that, as both n and ρ decrease, the fitted distributions obtained from MM and ML estimation are, or effectively are, wrapped half-normal, whatever the degree of asymmetry of the parent population. This fact is most clearly reflected in the content of Table 6.10. As can be appreciated from that table, for small ρ–values the problem of boundary ML estimates of γ_1 is more acute than that of inadmissible MM estimates of γ_1, particularly when n is small. As both n and ρ increase, the frequency of inadmissible MM estimates for γ_1 outstrips the corresponding frequency of boundary ML estimates. Clearly, if the ML estimate of γ_1 is a boundary estimate it follows that those for μ_L and σ correspond to the best fitting wrapped half-normal distribution. It would appear reasonable to interpret an inadmissible MM estimate

Table 6.10 Performance measures for the MM and ML estimates of γ_1. The measures given are: mean; (mean squared error); {percentage of inadmissible or boundary estimates, respectively}; [percentage of samples for which MM estimate was inadmissible and ML estimate was a boundary estimate]. For each (n, ρ, λ) combination the measures were calculated using 3000 simulated samples of size n from the WSNC distribution with $\xi = 0$ and ρ and λ as specified.

		ρ	0.3		0.5		0.7		0.9	
λ	n		MM	ML	MM	ML	MM	ML	MM	ML
0	20		−0.0036	−0.0016	0.0028	0.0083	−0.0008	−0.0087	−0.0083	−0.0028
			(0.6786)	(0.9463)	(0.5425)	(0.8320)	(0.4442)	(0.6143)	(0.4327)	(0.5763)
			{42.57}	{94.23}	{27.67}	{78.90}	{19.67}	{52.63}	{22.87}	{48.10}
			[41.77]		[24.27]		[13.63]		[14.40]	
	50		−0.0054	0.0162	0.0046	−0.0025	0.0032	−0.0028	−0.0267	−0.0143
			(0.6130)	(0.8759)	(0.3661)	(0.4877)	(0.2404)	(0.2147)	(0.2514)	(0.1854)
			{28.20}	{81.57}	{8.83}	{30.87}	{2.47}	{7.40}	{6.63}	{5.07}
			[25.63]		[3.93]		[0.53]		[0.93]	
	100		−0.0181	−0.0143	−0.0099	−0.0101	0.0021	0.0034	0.0045	0.0038
			(0.4967)	(0.7015)	(0.2289)	(0.2308)	(0.1368)	(0.0865)	(0.1412)	(0.0727)
			{13.10}	{52.60}	{1.40}	{6.00}	{0.40}	{0.23}	{0.90}	{0.20}
			[8.90]		[0.20]		[0]		[0]	
2	20		−0.4022	−0.4165	−0.2106	−0.2054	−0.0716	−0.0716	−0.0231	−0.0323
			(0.6194)	(0.6200)	(0.4652)	(0.4971)	(0.3284)	(0.3679)	(0.3133)	(0.3423)
			{45.37}	{95.60}	{29.03}	{79.47}	{26.47}	{59.23}	{35.27}	{55.27}
			[44.57]		[25.83]		[19.10]		[23.17]	
	50		−0.3219	−0.3312	−0.1204	−0.1100	0.0192	−0.0167	0.1003	−0.0101
			(0.5886)	(0.5820)	(0.3057)	(0.3423)	(0.1783)	(0.1568)	(0.1906)	(0.1328)
			{27.33}	{80.93}	{11.73}	{36.87}	{11.73}	{12.50}	{25.40}	{9.27}
			[24.27]		[6.13]		[2.93]		[4.40]	
	100		−0.2771	−0.2940	−0.0575	−0.0695	0.0561	−0.0177	0.1833	−0.0010
			(0.5206)	(0.5255)	(0.1905)	(0.1792)	(0.0991)	(0.0659)	(0.1336)	(0.0546)
			{13.60}	{55.33}	{4.97}	{8.33}	{4.67}	{0.87}	{21.47}	{0.73}
			[9.87]		[0.93]		[0.17]		[0.43]	
5	20		−0.6541	−0.6801	−0.3201	−0.3016	−0.1354	−0.0397	−0.0948	−0.0688
			(0.7047)	(0.4377)	(0.3979)	(0.2541)	(0.1929)	(0.1496)	(0.1704)	(0.1240)
			{48.43}	{95.77}	{41.07}	{85.27}	{48.50}	{66.27}	{59.77}	{77.53}
			[47.37]		[37.00]		[36.40]		[49.17]	
	50		−0.4131	−0.4196	−0.1351	−0.0751	−0.0154	−0.0122	0.0569	−0.0007
			(0.5190)	(0.3736)	(0.1601)	(0.1585)	(0.0599)	(0.0427)	(0.0419)	(0.0385)
			{37.73}	{81.53}	{36.07}	{44.77}	{45.90}	{35.77}	{73.23}	{36.17}
			[33.03]		[19.73]		[19.60]		[28.53]	
	100		−0.2546	−0.2429	−0.0486	−0.0300	0.0426	−0.0042	0.1067	0.0055
			(0.3400)	(0.2574)	(0.0640)	(0.0415)	(0.0248)	(0.0159)	(0.0231)	(0.0140)
			{29.93}	{56.07}	{34.83}	{20.20}	{46.63}	{9.97}	{81.20}	{8.03}
			[19.70]		[9.10]		[5.27]		[7.07]	

(continued overleaf)

Table 6.10 (Continued)

		ρ	0.3		0.5		0.7		0.9	
λ	n		MM	ML	MM	ML	MM	ML	MM	ML
20	20		−0.6463	−0.5991	−0.3379	−0.2041	−0.1698	−0.0411	−0.1570	−0.0631
			(0.6539)	(0.3186)	(0.3458)	(0.1284)	(0.1384)	(0.0500)	(0.1265)	(0.0492)
			{52.60}	{96.77}	{48.97}	{93.60}	{61.00}	{92.07}	{67.73}	{94.40}
			[51.83]		[47.40]		[57.80]		[65.47]	
	50		−0.3609	−0.2571	−0.1563	−0.0377	−0.0703	−0.0050	−0.0279	−0.0034
			(0.4119)	(0.1940)	(0.1093)	(0.0472)	(0.0344)	(0.0065)	(0.0163)	(0.0052)
			{49.50}	{90.30}	{52.30}	{80.97}	{65.63}	{87.50}	{85.83}	{87.73}
			[46.03]		[44.83]		[58.03]		[75.53]	
	100		−0.1741	−0.0790	−0.0825	−0.0106	−0.0260	0.0001	−0.0009	−0.0003
			(0.1666)	(0.0557)	(0.0356)	(0.0063)	(0.0074)	(0.0007)	(0.0025)	(0.0007)
			{53.17}	{82.20}	{56.23}	{63.13}	{72.83}	{65.53}	{93.77}	{64.13}
			[45.80]		[36.83]		[47.67]		[60.00]	

of γ_1 as indicating that the parent population is half-normal. However, the associated MM estimates of μ_L and σ will not generally be equal to the MM estimates obtained from fitting a wrapped half-normal distribution directly.

The tendencies of the two types of estimation to lead to solutions corresponding to wrapped half-normal distributions are reflected strongly in the MSEs of the individual estimates. When the underlying distribution is close to symmetric and highly dispersed, the MM estimates are competitive with, if not superior to, their ML counterparts. However, when the parent population is highly skew, or moderately skew but concentrated, the estimates obtained from ML estimation are superior.

6.7 Tests for Limiting Cases

As mentioned in Section 6.3.1, there are five limiting cases of the WSNC distribution. These are the: i) point, ii) uniform, iii) wrapped normal, iv) wrapped positive half-normal and v) wrapped negative half-normal, distributions. These five cases warrant special consideration as they can be expressed in terms of fewer parameters than the three of the WSNC class. Case i) is pathological and is easily identified. Tests for uniformity are legion; see, for example, Mardia & Jupp (1999, Section 6.3). Here we propose new procedures for testing for departures within the WSNC class from the other three limiting cases of the distribution. All three

procedures are based on the large-sample results for the sampling distribution of $\bar{b}_2$ established in Section 3.3 of Chapter 3.

6.7.1 A Large-sample Test for an Underlying Wrapped Normal Distribution

6.7.1.1 Derivation of the Test

Using the results in Equations (3.3.1) and (3.3.2), the relations (6.4.3) and (6.4.4) and the identities in (6.4.7), the sampling distribution of $\bar{b}_2$ under the null hypothesis of an underlying wrapped normal, i.e. a $\mathrm{WSNC_D}(\xi, \eta, 0)$, distribution is asymptotically normal with mean 0 and variance

$$\mathrm{var}_0\left(\bar{b}_2\right) = \frac{1}{n}\left\{\tfrac{1}{2}\left(1-\rho^{16}\right)+2\rho^4\left(\rho^8-\rho^6+\rho^2-1\right)\right\}$$

with $\rho = e^{-\frac{1}{2}\eta^2}$. Estimating ρ using $\bar{R}$, a moment based estimate of $\mathrm{var}_0\left(\bar{b}_2\right)$ is given by

$$\hat{\mathrm{var}}_0\left(\bar{b}_2\right) = \frac{1}{n}\left\{\tfrac{1}{2}\left(1-\bar{R}^{16}\right)+2\bar{R}^4\left(\bar{R}^8-\bar{R}^6+\bar{R}^2-1\right)\right\}.$$

Thus, a large-sample test for an underlying wrapped normal distribution can be based on the statistic

$$\frac{\bar{b}_2}{\left\{\hat{\mathrm{var}}_0\left(\bar{b}_2\right)\right\}^{1/2}}. \tag{6.7.1}$$

Large absolute values of (6.7.1) compared with the quantiles of the standard normal distribution lead to the rejection of the null hypothesis of an underlying wrapped normal distribution in favour of some skew alternative within the WSNC class.

6.7.1.2 The Asymptotic Power of the Test

For data sampled from a skew member of the WSNC class, the asymptotic distribution of $\bar{b}_2$ is again normal but with

$$E_1\left(\bar{b}_2\right) = \bar{\beta}_2 + \frac{1}{n}\left(-\frac{\bar{\beta}_3}{\rho}-\frac{\bar{\beta}_2}{\rho^2}+\frac{2\bar{\alpha}_2\bar{\beta}_2}{\rho^4}\right) \tag{6.7.2}$$

and

$$\mathrm{var}_1\left(\bar{b}_2\right) = \frac{1}{n}\left[\frac{1-\bar{\alpha}_4}{2}-2\bar{\alpha}_2-\bar{\beta}_2^2+\frac{2\bar{\alpha}_2}{\rho}\left\{\bar{\alpha}_3+\frac{\bar{\alpha}_2\left(1-\bar{\alpha}_2\right)}{\rho}\right\}\right]. \tag{6.7.3}$$

For a specific $\mathrm{WSNC_D}(\xi,\eta,\lambda)$ distribution, the values of (6.7.2) and (6.7.3) can be calculated using (6.4.3), (6.4.4) and (6.4.7). Given these asymptotic results, together with those for $\bar{b}_2$ under the null hypothesis, the asymptotic power of the test for a significance level of $100\alpha\,\%$ is

$$1-\Phi\left[\frac{z_{\alpha/2}\left\{\mathrm{var}_0\left(\bar{b}_2\right)\right\}^{1/2}-E_1\left(\bar{b}_2\right)}{\left\{\mathrm{var}_1\left(\bar{b}_2\right)\right\}^{1/2}}\right]+\Phi\left[\frac{-z_{\alpha/2}\left\{\mathrm{var}_0\left(\bar{b}_2\right)\right\}^{1/2}-E_1\left(\bar{b}_2\right)}{\left\{\mathrm{var}_1\left(\bar{b}_2\right)\right\}^{1/2}}\right]. \quad (6.7.4)$$

6.7.1.3 Empirical Investigation of the Operating Characteristics of the Test

In order to investigate the ability of the test to maintain the nominal significance level, and its power against skew alternatives from the WSNC class, we performed a further simulation experiment. These two fundamental operating characteristics of the test were explored using simulated samples of size, n, equal to 20, 30, 50, 100, 200 and 500, drawn from WSNC distributions with λ-values of 0, 2, 5 and ∞, and ρ-values of 0.1(0.05)0.95. For each (n, ρ, λ) combination we simulated 5000 samples of size n from the WSNC distribution with mean direction 0, mean resultant length ρ and skewness parameter λ. Each empirical size, or power, value, as appropriate, was established by performing the test for each of the 5000 samples concerned and noting the proportion of such samples for which the null hypothesis was rejected. The nominal significance levels investigated were 10%, 5% and 1%, although here we present results for just the 5% level.

The results for the test's ability to maintain the nominal significance level are represented graphically in Figure 6.11. Those for the sample sizes of 30 and 200 have been omitted from the figure in order to improve its clarity. We note that the standard error of each empirical result represented in this figure is approximately 0.003. As is evident from a consideration of the content of this figure, the test maintains the nominal significance level very well indeed, even for a sample size as small as 20.

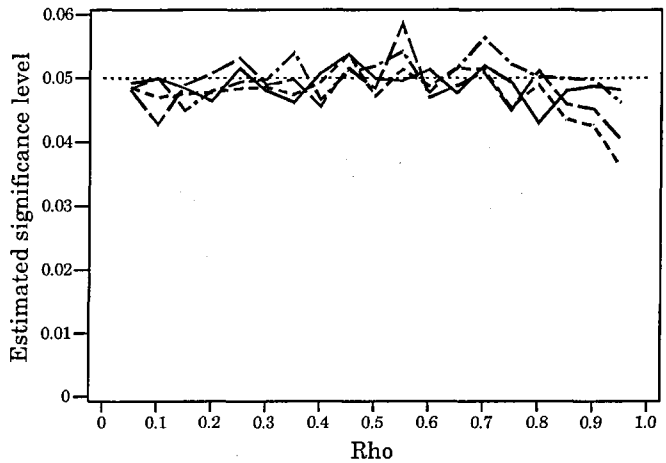

Figure 6.11 Estimated size of the test for the null hypothesis of an underlying wrapped normal distribution. Sample sizes represented are: - - - - - (n = 20); — — — — (n = 50); — · — (n = 100); ——— (n = 500). The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$.

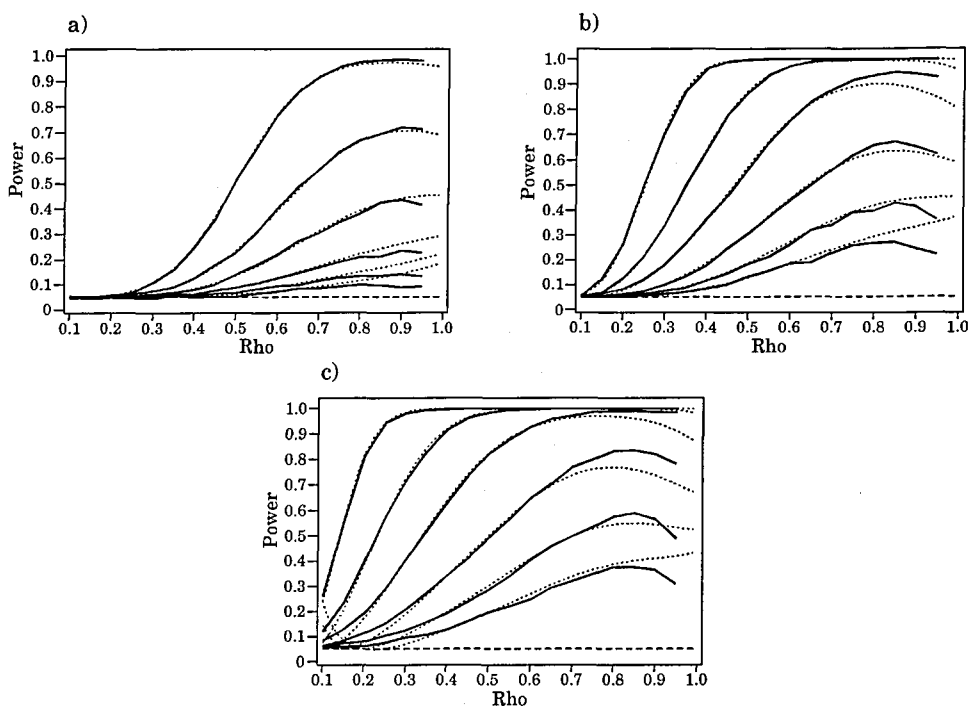

Figure 6.12 Theoretical asymptotic power (· · · · · ·) and empirical power (———) of the test for the null hypothesis of an underlying wrapped normal distribution when the parent population is WSNC with: a) $\lambda = 2$, b) $\lambda = 5$, c) $\lambda = \infty$. The six curves of each type correspond to sample sizes of 20, 30, 50, 100, 200 and 500, the power increasing with sample size. The dashed horizontal line delimits the nominal significance level of $\alpha = 0.05$.

Figure 6.12 provides a graphical representation of the empirical results obtained for the power of the test for the samples drawn from WSNC distributions with λ equal to 2, 5 and ∞. For comparative purposes, we have also included in the three plots making up this figure the corresponding theoretical asymptotic power functions calculated using (6.7.4). In general, the agreement between the empirical and theoretical results is very good, the largest differences between the two occurring for highly concentrated, and highly skew and dispersed, cases of the WSNC distribution. As is to be expected, such disparities tend to be most pronounced for samples of small size. A comparison of the results portrayed in this figure with their counterparts in Figure 4.4 provides an assessment of the gain in power of the present test over that of the omnibus test for symmetry introduced in Section 4.3 of Chapter 4.

6.7.2 A Large-sample Test for an Underlying Wrapped Half-normal Distribution

6.7.2.1 Derivation of the Test

In the definition of a test for an underlying wrapped (positive) half-normal distribution based on the asymptotic results for the sampling distribution of $\bar{b}_2$ introduced in Chapter 3, one first needs to specify how the direct parameter η might be estimated. From (6.4.3), when $\delta = 1$, we have

$$\rho^2 = e^{-\eta^2}\{1+\Im^2(\eta)\}.$$

Estimating ρ using the moment estimate $\bar{R}$, an estimate, $\tilde{\eta}$, of η is provided by the solution to

$$\bar{R}^2 - e^{-\eta^2}\{1+\Im^2(\eta)\} = 0, \tag{6.7.5}$$

which must be solved using numerical methods.

An estimate of the theoretical value of $\bar{\beta}_2$ for an assumed underlying wrapped (positive) half-normal distribution, $\bar{\beta}_2(\tilde{\eta})$, say, can then be calculated from (6.4.4) with $\delta = 1$ and η replaced by $\tilde{\eta}$. Similarly, an estimate of the large-sample variance of $\bar{b}_2$ for an underlying wrapped (positive) half-normal distribution, $\hat{\text{var}}_0(\bar{b}_2)$, say, can be obtained using (6.7.3) together with (6.4.3), (6.4.4) and (6.4.7), with δ set equal to 1, and η equal to $\tilde{\eta}$.

A large-sample test for an underlying wrapped (positive) half-normal distribution can then be based on the statistic

$$\frac{\bar{b}_2 - \bar{\beta}_2(\tilde{\eta})}{\left\{\hat{\mathrm{var}}_0\left(\bar{b}_2\right)\right\}^{1/2}}. \tag{6.7.6}$$

Large positive values of (6.7.6) compared with the upper quantiles of the standard normal distribution lead to the rejection of the hypothesis of an underlying wrapped (positive) half-normal distribution in favour of some less positively skew alternative within the WSNC class.

For an equivalent test for an underlying wrapped negative half-normal distribution, the test statistic becomes

$$\frac{\bar{b}_2 + \bar{\beta}_2(\tilde{\eta})}{\left\{\hat{\mathrm{var}}_0\left(\bar{b}_2\right)\right\}^{1/2}}. \tag{6.7.7}$$

Here, large negative values of (6.7.7) compared with the lower quantiles of the standard normal distribution lead to the rejection of the hypothesis of an underlying wrapped negative half-normal distribution in favour of some less negatively skew alternative within the WSNC class.

6.7.2.2 The Asymptotic Power of the Test

Using the asymptotic results of Chapter 3 once more, under the null hypothesis of an underlying wrapped (positive) half-normal distribution, equivalent to a $\mathrm{WSNC}_{\mathrm{D}}(\xi,\eta,\infty)$ distribution, the asymptotic distribution of $\bar{b}_2$ is normal with mean

$$E_0\left(\bar{b}_2\right) = \bar{\beta}_2 + \frac{1}{n}\left(-\frac{\bar{\beta}_3}{\rho} - \frac{\bar{\beta}_2}{\rho^2} + \frac{2\bar{\alpha}_2\bar{\beta}_2}{\rho^4}\right) \tag{6.7.8}$$

and variance

$$\mathrm{var}_0\left(\bar{b}_2\right) = \frac{1}{n}\left[\frac{1-\bar{\alpha}_4}{2} - 2\bar{\alpha}_2 - \bar{\beta}_2^2 + \frac{2\bar{\alpha}_2}{\rho}\left\{\bar{\alpha}_3 + \frac{\bar{\alpha}_2\left(1-\bar{\alpha}_2\right)}{\rho}\right\}\right]. \tag{6.7.9}$$

The mean resultant length and the central trigonometric moments appearing in (6.7.8) and (6.7.9) can be calculated using (6.4.3), (6.4.4) and (6.4.7) with δ set equal to 1.

Under the alternative hypothesis of an underlying WSNC distribution with $\lambda \neq \infty$, the asymptotic distribution of $\bar{b}_2$ is normal with mean $E_1\left(\bar{b}_2\right)$ and

variance $\text{var}_1\left(\bar{b}_2\right)$. The values of $E_1\left(\bar{b}_2\right)$ and $\text{var}_1\left(\bar{b}_2\right)$ can be obtained in the same way as described for $E_0\left(\bar{b}_2\right)$ and $\text{var}_0\left(\bar{b}_2\right)$, the only difference being that the equivalent value of δ corresponding to the specified value of λ should be used instead of $\delta = 1$. It follows that the asymptotic power of the test for a significance level of $100\alpha\,\%$ is given by

$$1-\Phi\left[\frac{z_\alpha\left\{\text{var}_0\left(\bar{b}_2\right)\right\}^{1/2}-E_1\left(\bar{b}_2\right)+E_0\left(\bar{b}_2\right)}{\left\{\text{var}_1\left(\bar{b}_2\right)\right\}^{1/2}}\right]. \qquad (6.7.10)$$

6.7.2.3 Empirical Investigation of the Operating Characteristics of the Test

As in Section 6.7.1.3, we used simulation to investigate the present test's ability to maintain the nominal significance level and its power. This investigation made use of simulated samples of size, n, equal to 20, 30, 50, 100, 200 and 500 drawn from WSNC distributions with λ-values of ∞, 5, 2, 0, –2, –5 and $-\infty$, and ρ-values of 0.3(0.05)0.95. The lower value of ρ was set at 0.3 so as to avoid potential overflow problems in FORTRAN during the computation of the values of the function $\Im(\cdot)$ required in the calculation of the central trigonometric moments appearing in (6.4.7). The empirical size, or power, value corresponding to a particular (n, ρ, λ) combination was computed using 5000 samples of size n simulated from the WSNC distribution with mean direction 0, mean resultant length ρ and skewness parameter λ. Again, the nominal significance levels investigated were 10%, 5% and 1%. The results presented below are those obtained for the nominal significance level of 5%.

Figure 6.13 provides a graphical representation of the results obtained for the empirical size of the test, calculated using the samples simulated from the wrapped (positive) half-normal distribution, i.e. the $\text{WSNC}_{\text{D}}(\xi,\eta,\infty)$ distribution. The standard error of any empirical result portrayed in this figure is, at most, 0.0053. From this figure we see that, for ρ-values less than 0.6, the test is non-conservative. However, as the sample size increases, the ability of the test to maintain the nominal significance level improves. For ρ-values in excess of 0.7 the test is conservative, its conservatism increasing with the concentration of the parent population and the sample size.

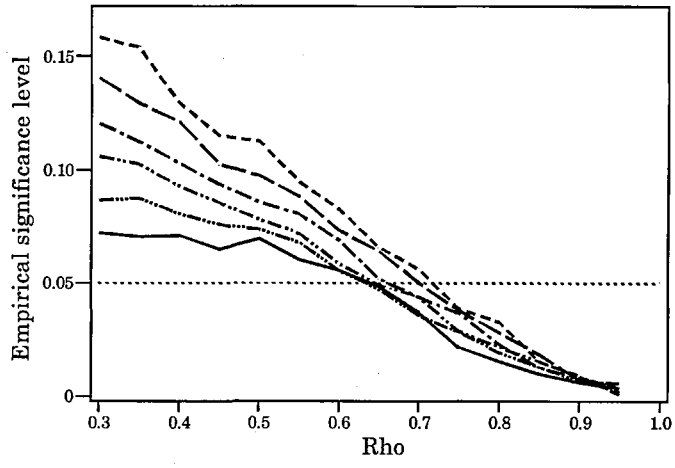

Figure 6.13 Estimated size of the test for the null hypothesis of an underlying wrapped (positive) half-normal distribution. Sample sizes represented are: – – – – (n = 20); — — — (n = 30); — · — (n = 50); — · · — (n = 100); — · · · — (n = 200); ——— (n = 500). The dotted horizontal line delimits the nominal significance level of $\alpha = 0.05$.

The empirical results for the power of the test are portrayed in the six plots making up Figure 6.14. Also included in these plots are the corresponding theoretical asymptotic power functions calculated from (6.7.10). As is to be expected, the agreement between the two sets of results improves with increasing sample size. Nevertheless, in general, the theoretical asymptotic power tends to underestimate the true power of the test. We see that the power of the test generally increases with increasing ρ, before reaching a maximum and subsequently decreasing as ρ increases yet further. This drop off in power as ρ tends to 1 is consistent with the corresponding increasing conservatism of the test noted previously.

6.7.2.4 A Monte Carlo Variant of the Test

As observed in the previous section, the disparities between the theoretical and empirical power results for the large-sample test are often sizeable. Such disparities are perhaps to be expected for two principal reasons. Firstly, there is the potential source of bias introduced by the need to estimate η. Secondly, as is evident from Figure 6.6f, the normal distribution does not provide a good approximation to the sampling distribution of $\bar{b}_2$ for small sized samples drawn from highly skew parent WSNC populations.

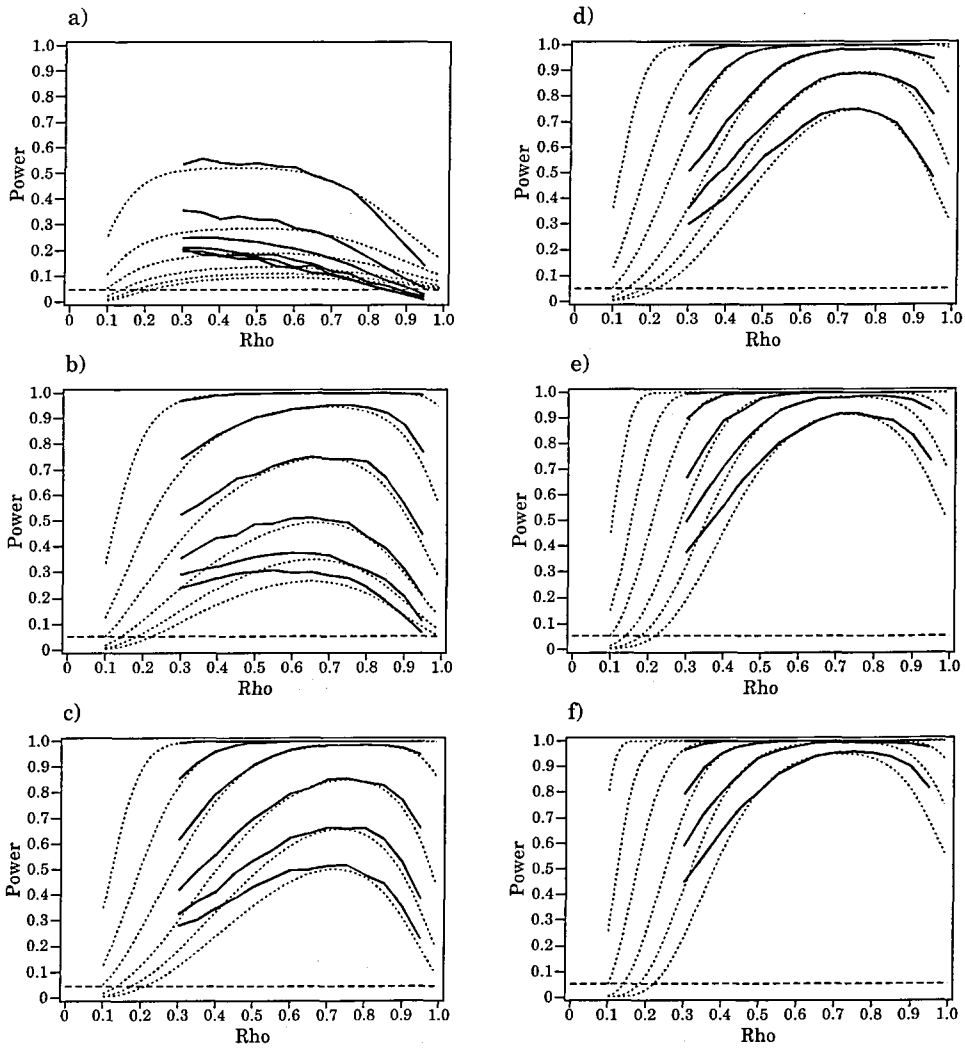

Figure 6.14 Theoretical asymptotic power (· · · · ·) and empirical power (———) of the test for the null hypothesis of an underlying wrapped (positive) half-normal distribution when the parent population is WSNC with: a) $\lambda = 5$, b) $\lambda = 2$, c) $\lambda = 0$, d) $\lambda = -2$, e) $\lambda = -5$, f) $\lambda = -\infty$. The six curves of each type correspond to sample sizes of 20, 30, 50, 100, 200 and 500, the power increasing with sample size. The dashed horizontal line delimits the nominal significance level of $\alpha = 0.05$.

Instead of using a test based on a large-sample normal approximation to the sampling distribution of (6.7.6) we might choose to use a Monte Carlo version of the test which approximates the sampling distribution of an equivalent test statistic empirically. Such a variant of the test can be based upon the rank associated with the value of $\bar{b}_2$ for the original data when ordered amongst the values of the same statistic calculated for simulated samples of the same size from a $\mathrm{WSNC}_{\mathrm{D}}(0, \eta, \infty)$ distribution with η estimated using the solution to (6.7.5). As this computer

intensive approach to testing obviates the need for a normal approximation, one might expect this variant of the test to be capable of maintaining the nominal significance level better. However, it is not obvious whether its power will be higher than that of the original large-sample version of the test. Also, although the use of simulation circumvents potential problems associated with the inadequacy of the normal approximation, the need to estimate η persists.

6.8 An Illustrative Example

6.8.1 Introduction to the Data

The data set we use to illustrate our methodology was introduced in the ornithological literature by Bruderer & Jenni (1990), and consists of the "headings" of 1827 birds recorded at an observational post near Stuttgart during the autumnal migratory period of 1987. Here, the term "heading" refers to the direction, measured in a clockwise direction from magnetic North, of a bird's body during flight. The data in question are listed in their entirety in Appendix 1 and summarized in the form of a frequency distribution in Table 6.11. A linear histogram of the data, converted to radians in $[0, 2\pi)$ and grouped using class intervals of width equivalent to 10°, is given in Figure 6.15. From a consideration of this linear histogram it is evident that the distribution is negatively skew with a modal direction of around 4 radians, or 230°.

Table 6.11 Frequency distribution of the headings of 1827 birds recorded during the autumnal migratory period of 1987.

Heading (degrees)	Frequency	Heading (degrees)	Frequency	Heading (degrees)	Frequency
[0–10)	2	[120–130)	10	[240–250)	169
[10–20)	4	[130–140)	15	[250–260)	163
[20–30)	4	[140–150)	25	[260–270)	116
[30–40)	3	[150–160)	35	[270–280)	79
[40–50)	3	[160–170)	54	[280–290)	53
[50–60)	4	[170–180)	68	[290–300)	25
[60–70)	8	[180–190)	84	[300–310)	9
[70–80)	6	[190–200)	120	[310–320)	12
[80–90)	8	[200–210)	162	[320–330)	6
[90–100)	8	[210–220)	161	[330–340)	6
[100–110)	14	[220–230)	168	[340–350)	6
[110–120)	12	[230–240)	200	[350–360)	5

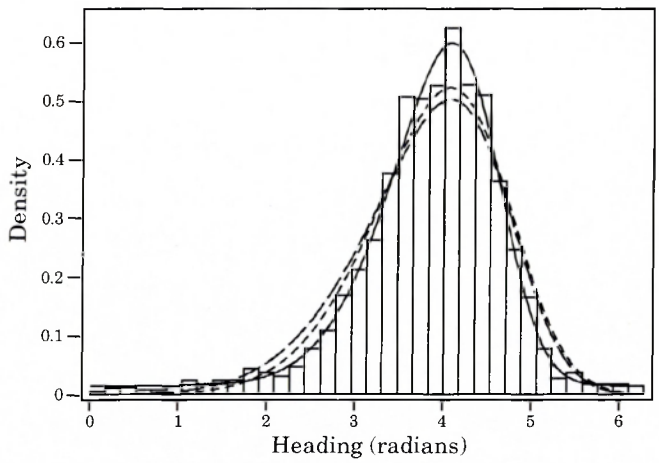

Figure 6.15 Histogram of the bird-flight headings together with fitted densities: – – –, MM solution $WSNC_D(4.66, 1.10, -1.79)$; — —, ML solution $WSNC_D(4.70, 1.21, -2.22)$; ——, ML solution $UWSNC_D(0.10, 4.56, 0.92, -2.07)$.

6.8.2 Fit of the WSNC Distribution

When applied, the large sample tests of Section 6.7 emphatically rejected the wrapped normal, wrapped positive half-normal and wrapped negative half-normal distributions as potential models for the underlying distribution, their respective *p*-values all being effectively equal to 0.

We therefore explored the fit of the wrapped skew-normal distribution on the circle. Proceeding as described in Sections 6.4.2 and 6.4.3, the MM solution, obtained by estimating the circular parameters of the distribution and then transforming to the direct parameters, was found to be $WSNC_D(4.66, 1.10, -1.79)$. For the ML estimation of the parameters we used the Nelder-Mead simplex to optimize the direct parameter representation of the log-likelihood derived from (6.3.1). The use of this representation of the log-likelihood was not considered problematic for these data given the prior rejection of the hypothesis of an underlying wrapped normal distribution. The fitted distribution corresponding to the ML solution was $WSNC_D(4.70, 1.21, -2.22)$, with an associated value of the log-likelihood of –2202.06. In Figure 6.15, the densities for these two solutions have been overlaid upon the linear histogram of the data. Visually, the densities are very similar. However, it is clear that neither density provides a particularly good fit to the distribution of the data. Specifically, both fitted densities are incapable of simultaneously modelling the peakedness and long 'tails' evident in the empirical distribution. These visual impressions are supported by the results from chi-squared goodness-of-fit analyses. The value of the chi-squared statistic for the MM

solution, based on 26 class intervals with expected frequencies in excess of 5, was 244.32. The equivalent value of the test statistic for the ML solution was 152.94. Comparing these two values with the quantiles of the chi-squared distribution on 22 degrees of freedom, the associated p-values are both effectively equal to 0. Hence the WSNC distribution is emphatically rejected as a suitable model for these data.

6.8.3 Fit of a Uniform and WSNC Mixture Model

In our search for a plausible model which could adequately model the data, we pursued the logic of the following argument. Given the dominant direction of migration and the angular spread of the data, it would appear improbable that the directions making up the distribution's tails correspond to birds moving in the general direction of migration. It seems more likely that these directions correspond to birds pursuing other needs of an avian existence, such as tracking down the next meal or pursuing a potential mate. Moreover, as we have no reason to assume that the directions followed by such birds are restricted exclusively to the tails of the empirical distribution, a circular uniform distribution suggests itself as a potential model for the headings of this background of presumably non-migrating birds. As, overall, the empirical distribution is asymmetric, a skew member of the parsimonious WSNC class would appear to be a reasonable model for the data making up the remainder of the empirical distribution. Viewed in this way, the form taken by the empirical distribution is a consequence of the mixing of data from two distinct populations. As there is no additional information which might be used to classify the individuals as belonging to one or other of the two component distributions, we explored the fit of the model with density

$$f(\theta;p,\xi,\eta,\lambda)=\frac{p}{2\pi}+(1-p)\frac{2}{\eta}\sum_{r=-\infty}^{\infty}\phi\left(\frac{\theta+2\pi r-\xi}{\eta}\right)\Phi\left\{\lambda\left(\frac{\theta+2\pi r-\xi}{\eta}\right)\right\}, \quad (6.8.1)$$

corresponding to a mixture distribution with circular uniform and WSNC components. We denote the distribution having this density as $\mathrm{UWSNC_D}(p,\xi,\eta,\lambda)$, where p denotes the mixing probability and, once more, the subindex D refers to the direct parametrization.

The ML solution obtained from maximizing the log-likelihood derived from (6.8.1) was $\mathrm{UWSNC_D}$(0.10, 4.56, 0.92, –2.07), the associated value of the log-likelihood being –2128.03. The density corresponding to this ML solution is also displayed in Figure 6.15. A comparison of this density with the histogram and the other two fitted densities represented in that figure provides us with the visual

impression of an overall improvement in fit. The results obtained from the use of a generalized likelihood ratio test also support this impression of an improvement in fit over that of the best fitting WSNC distribution. Comparing $-2(-2202.06 + 2128.03) = 148.06$ with the quantiles of the chi-squared distribution with 1 degree of freedom, the improvement in fit is highly significant, the p-value of the test effectively being equal to 0. We also conducted a chi-squared goodness-of-fit test in order to investigate the overall fit of the mixture model. Using 33 class intervals with expected frequencies in excess of 5, the value of the chi-squared test statistic was found to be 34.03 on 28 degrees of freedom. The p-value of around 0.2 associated with this result confirms our previous impressions regarding the fit of the mixture model.

In conclusion, the proposed mixture model provides a much improved fit to the empirical distribution and a plausible behavioural model for the data containing just one additional parameter to those of the WSNC class. According to the fitted density, 10% of the headings correspond to birds making up the circular uniform background, and the remaining 90% to birds whose headings are distributed according to the $\text{WSNC}_{\text{D}}(4.56, 0.92, -2.07)$ distribution.

6.9 Summary and Directions for Future Research

We close the present chapter with a summary of the main points that have so far been addressed within it and a discussion of some potential lines for future related research.

6.9.1 Summary

Chapters 4 and 5 were dedicated to testing for circular reflective symmetry. In this chapter we have addressed the issue as to how one might proceed if the circular data under consideration are identified as being drawn from an underlying skew distribution. In Section 6.2 we briefly reviewed the asymmetric models which have been proposed in the literature for use with circular data, outlining in passing some shortcomings of two models proposed by Papakonstantinou (1979) and Batschelet (1981).

The remainder of the chapter was devoted to a detailed consideration of the wrapped skew-normal distribution on the circle. In Section 6.3 we provided a definition of the distribution and derived its fundamental properties such as its characteristic function and trigonometric moments.

Proceeding to a consideration of issues of inference for the distribution, in Section 6.4 we described how the parameters of the distribution might be estimated using the method of moments. Our approach hinged upon the use of the so-called circular parametrization of the distribution involving the mean direction, mean resultant length and second central sine moment. These were the circular measures for which large-sample inferential procedures were derived in Chapter 3. In Section 6.4.2 we gave the relations between the circular and direct parameters, and described how the general results of Chapter 3 might be used to conduct inference for the circular parameters of the distribution. Various sampling properties of the method of moments were investigated in Section 6.4.3. Specifically, we presented empirical results obtained from a simulation experiment designed to explore the shapes of the sampling distributions of the estimates, the validity of the theoretical asymptotic results for the bias and variance of the estimates, and the efficacy of bias-correction. These empirical results provided evidence of the generally regular forms assumed by the sampling distributions of the estimates, and the wide-ranging validity of the theoretical asymptotic bias and variance results of Chapter 3 when applied to the estimation of the circular parameters of the WSNC distribution. On the other hand, the results obtained for the general bias-corrected estimates introduced in Chapter 3 led us to the conclusion that their use in the estimation of the circular parameters of the WSNC distribution cannot be recommended.

Although the circular parameters of the WSNC distribution may well be of interest in their own right, in other contexts the direct, or centred, parameters of the distribution will be of greater interest. In Section 6.4.4 we considered the forms assumed by the sampling distributions of the estimates of the direct and centred parameters obtained by transforming the method of moments estimates of the circular parameters. In comparison with the sampling distributions of the latter, those for the corresponding estimates of the direct and centred parameters were classified as being, in general, far less regular.

Section 6.5 was dedicated to likelihood based methods of estimation. We proposed an approach to maximizing the log-likelihood function, expressed in terms of the centred parameters, based on a grid based search used in conjunction with the Nelder-Mead simplex. We also gave details of how estimates of the direct or circular parameters can be obtained by transforming those for the circular parameters. In general, the sampling distributions of the resulting estimates were found to be unappealing.

In Section 6.6 we presented empirical results derived from a detailed simulation based investigation of the sampling properties of the method of moments and maximum likelihood estimates. Generally, the estimates obtained using the maximum likelihood method were identified as having superior properties, apart from when the underlying distribution is close to symmetric and highly dispersed, in which case the method of moments estimates can be superior.

Tests for the limiting cases of the wrapped normal, wrapped positive half-normal and wrapped negative half-normal distributions were proposed in Section 6.7. The derivation of all three test procedures drew on the large-sample results for $\bar{b}_2$ given in Chapter 3. In Sections 6.7.1.2 and 6.7.2.2 we gave results for the theoretical asymptotic power of two of the tests, and in Sections 6.7.1.3 and 6.7.2.3 presented empirical results for two of their primary operating characteristics. The test for an underlying wrapped normal distribution was found to maintain the nominal significance level very well, even for sample sizes as small as 20. Also, the agreement between the theoretical asymptotic power and the true empirical power of the test was identified as being very close. Our large-sample test for an underlying (positive) half-normal distribution is somewhat more involved as it requires the estimation of the dispersion parameter η. We found that this test does not hold the nominal significance level well, and that the agreement between its theoretical asymptotic power and true empirical power is not particularly good for certain sample size and mean resultant length combinations. With a view to improving the operating characteristics of this test, in Section 6.7.2.4 we proposed a Monte Carlo variant of it.

In order to illustrate certain aspects of the methodology elaborated throughout the chapter, in Section 6.8 we presented an analysis of a large data set consisting of the headings of migrating birds. Although, on its own, the WSNC distribution was found not to adequately model these data, it did provide the major component of a parsimonious mixture model which fitted the data well.

6.9.2 Directions for Future Research

In this chapter we have concentrated principally on the basic properties of the WSNC class, the point estimation of its parameters using the method of moments and maximum likelihood approaches, and tests for three of its limiting cases. Clearly, the contributions made within the chapter address only some of the many issues of potential interest regarding the WSNC class. It would, for instance, be

interesting to compare the performance of MM and ML based estimation with that of other approaches. In this context, an extension of the work of Liseo (1990) on Bayesian inference to deal with data distributed on the circle suggests itself as a potentially fruitful line of investigation.

The method of moments based tests for the limiting cases of the WSNC class discussed in Section 6.7 are appealing as, being based on $\bar{b}_2$, they are relatively simple to perform. This is particularly true for the test for an underlying wrapped normal distribution. As we saw in Section 6.7.1.3, this test also has good operating characteristics. Those for the test for an underlying wrapped half-normal distribution are not as impressive. It would therefore be interesting to compare the operating characteristics of the large-sample version of this latter test with those of the Monte Carlo variant of it described in Section 6.7.2.4. Alternatively, rather than use a method of moments based test, another possibility would be to use generalized likelihood ratio procedures to test for the limiting cases of the class. One inconvenience of such tests is that they require specialist software in order to numerically optimize the log-likelihood function of an assumed underlying WSNC distribution. Another is that, when testing for an underlying wrapped half-normal distribution, the standard asymptotic properties of the generalized likelihood ratio statistic cannot be appealed to, a wrapped half-normal distribution corresponding to one on the boundary of the parameter space of the WSNC class. Given the form of the sampling distribution of the ML estimate of γ_1 displayed in Figure 6.9c, one might also envisage potential problems with applying standard theory to a test for an underlying wrapped normal distribution based on the generalized likelihood ratio statistic.

Similar comments apply regarding the use of standard likelihood theory in the construction of confidence sets for the parameters of the WSNC distribution. In contrast, the sampling distributions of the MM estimates are, as we have seen, particularly well behaved and the large-sample results of Chapter 3 can be used in conjunction with the assumption of an underlying WSNC distribution to easily derive confidence regions for the distribution's circular parameters. If required, confidence regions for the direct or centred parameters can be obtained by transforming those for the circular parameters.

In testing the goodness-of-fit of the WSNC distribution to the bird migration data of Section 6.8, we used the standard chi-squared procedure. Whilst this is certainly a relatively easy test to apply, its operating characteristics are known to

depend heavily on the choice of class intervals used in its application. In principle, it would be possible to develop variants of Kuiper's test or Watson's U^2 test for testing the goodness-of-fit of a WSNC distribution whose parameters have been estimated. We note that the only work published specifically on testing the fit of an assumed (linear) skew-normal distribution is that of Gupta & Chen (2001). These authors provide tables of the distribution function of the standard skew-normal distribution which can be used in conjunction with the usual chi-squared or Kolmogorov-Smirnov tests to test the fit of a specified standard skew-normal distribution.

Although, on its own, the WSNC distribution did not adequately model the bird migration data, it did contribute the major component of a mixture model found to fit the data well. This fact points up an important potential use of the WSNC distribution. Whilst, as, for example, McLachlan & Peel (2000, Chapter 1) show, asymmetry can be modelled using mixture models with symmetric component distributions, the use of WSNC distributions as components of finite mixture models raises the possibility of modelling skew distributed circular data more parsimoniously. Of course, this assumes we have no *a priori* reason for using symmetric component distributions in the mixture based modelling of asymmetry. Similar arguments apply to the potential role of the WSNC distribution in the modelling of multimodal circular data exhibiting asymmetry.

The primary focus of the current chapter has been the WSNC distribution, a distribution capable of modelling varying degrees of asymmetry. The methodology developed and illustrated in the preceding sections hopefully provides a strong case for the potential applicability of the WSNC distribution in the modelling of skew circular data. However, in Section 6.2 we referred to two families of distributions which, due to their greater flexibility, perhaps have even greater potential in this role. An ambitious line of future potential research would be to develop methods of inference for the wrapped Pearson and wrapped stable families.

Instead of assuming a given parametric form for an underlying distribution, an alternative modern approach to inference is that of empirical likelihood; see, for example, Owen (2001). The associated methodology draws heavily on other modern methods such as the bootstrap and kernel based density estimation. Although there now exists an extensive literature associated with empirical likelihood, we are aware of only one application of this methodology in the analysis of directional data. Fisher *et al.* (1996) combine the use of empirical likelihood and the bootstrap

to construct confidence regions for the mean direction associated with data distributed on the surface of the unit sphere.

Appendix 1 The Bird-flight Headings of Bruderer & Jenni (1990)

The headings, in degrees, are listed in increasing numerical order.

5	8	14	16	18	19	21	24	29	29	33	34	35	44	45	47	52	54
55	55	61	61	62	62	64	67	68	69	70	71	72	78	78	78	80	84
85	85	87	87	88	89	92	92	95	96	97	98	99	99	100	102	102	102
103	103	104	104	105	107	108	109	109	109	110	110	111	111	112	115	115	116
117	117	118	119	121	124	124	125	125	125	126	128	128	129	131	133	133	134
134	134	134	135	135	136	137	138	138	139	139	140	142	142	142	142	143	143
143	144	144	144	145	145	146	146	146	147	147	147	148	148	149	149	149	149
151	151	152	152	152	153	153	153	153	153	153	153	154	154	154	154	155	155
155	155	155	155	155	156	156	156	157	157	157	158	159	159	159	159	159	160
160	160	160	161	162	162	162	162	162	162	163	163	163	163	164	164	164	164
164	164	165	165	165	166	166	166	166	166	166	166	166	166	167	167	167	167
167	167	167	168	168	168	168	168	168	168	168	168	168	169	169	169	169	170
170	170	170	170	171	171	171	171	171	171	171	171	171	172	172	172	172	173
173	173	173	173	173	173	173	173	174	174	174	174	174	174	174	175	175	175
175	175	175	176	176	176	176	176	176	176	176	176	177	177	177	177	177	177
177	177	178	178	178	178	178	178	178	179	179	179	179	180	180	180	180	180
180	180	180	180	181	181	181	181	181	181	181	181	181	181	181	181	181	181
182	182	182	182	182	182	182	183	183	183	183	183	184	184	184	184	184	184
184	184	184	184	185	185	185	186	186	186	186	186	186	186	186	186	187	187
187	187	187	187	187	187	187	188	188	188	188	188	188	188	188	189	189	189
189	189	189	189	189	189	189	190	190	190	190	190	190	190	191	191	191	191
191	191	191	191	191	191	191	192	192	192	192	192	192	192	192	192	193	193
193	193	193	193	193	193	193	193	193	193	193	194	194	194	194	194	194	194
194	194	194	194	194	195	195	195	195	195	195	195	195	195	195	195	195	195
195	195	195	196	196	196	196	196	196	196	196	196	196	196	197	197	197	197
197	197	197	197	197	197	197	197	197	197	197	197	198	198	198	198	198	198
198	198	198	198	198	198	199	199	199	199	199	199	199	199	199	199	199	199
199	200	200	200	200	200	200	200	200	200	200	200	200	200	201	201	201	201
201	201	201	201	201	201	201	202	202	202	202	202	202	202	202	202	202	202
203	203	203	203	203	203	203	203	203	203	203	203	203	204	204	204	204	204
204	204	204	204	204	204	204	204	204	204	204	204	204	205	205	205	205	205
205	205	205	205	205	205	205	205	205	205	205	205	205	205	205	205	205	206
206	206	206	206	206	206	206	206	206	206	206	206	206	206	207	207	207	207
207	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207
208	208	208	208	208	208	208	208	208	208	208	208	208	208	208	208	209	209
209	209	209	209	209	209	209	209	209	209	209	209	209	209	209	209	209	209
209	210	210	210	210	210	210	210	210	210	210	210	210	210	210	210	211	211
211	211	211	211	211	212	212	212	212	212	212	212	212	212	212	212	212	212
212	213	213	213	213	213	213	213	213	213	213	213	213	213	213	213	213	213
213	213	214	214	214	214	214	214	214	214	214	214	214	214	214	215	215	215
215	215	215	215	215	215	215	215	215	215	215	216	216	216	216	216	216	216
216	216	216	216	216	216	216	216	216	216	216	216	216	216	216	216	216	216
216	216	216	217	217	217	217	217	217	217	217	217	217	217	217	218	218	218
218	218	218	218	218	218	218	218	218	218	218	218	218	218	218	218	218	218
219	219	219	219	219	219	219	219	219	219	219	219	219	219	219	219	219	219
220	220	220	220	220	220	220	220	220	220	220	220	220	220	220	220	220	221
221	221	221	221	221	221	221	221	221	221	221	221	222	222	222	222	222	222
222	222	222	222	222	222	222	222	222	222	222	222	222	223	223	223	223	223

223 223 223 223 223 223 223 223 223 224 224 224 224 224 224 224 224 224
224 224 224 224 224 224 224 225 225 225 225 225 225 225 225 225 225 225
225 225 225 225 225 225 225 226 226 226 226 226 226 226 226 226 226 226
226 226 226 226 226 226 226 226 226 226 226 226 226 226 227 227 227 227
227 227 227 227 227 227 227 227 227 228 228 228 228 228 228 228 228 228
228 228 228 228 228 228 228 228 229 229 229 229 229 229 229 229 229 229
229 229 229 229 229 229 230 230 230 230 230 230 230 230 230 230 230 230
230 230 230 230 230 230 230 230 230 230 230 230 230 230 230 231 231 231
231 231 231 231 231 231 231 231 231 231 232 232 232 232 232 232 232 232
232 232 232 232 232 232 233 233 233 233 233 233 233 233 233 233 233 233
233 233 233 233 233 233 233 233 234 234 234 234 234 234 234 234 234 234
234 234 234 234 234 234 234 234 234 234 235 235 235 235 235 235 235 235
235 235 235 235 235 235 236 236 236 236 236 236 236 236 236 236 236 236
236 236 236 236 236 236 236 236 236 236 236 236 236 236 236 237 237 237
237 237 237 237 237 237 237 237 237 237 237 237 237 237 237 237 237 237
237 237 237 237 238 238 238 238 238 238 238 238 238 238 238 238 238 238
238 238 238 238 239 239 239 239 239 239 239 239 239 239 239 239 239 239
239 239 239 239 239 239 239 239 240 240 240 240 240 240 240 240 240 240
240 240 240 240 240 240 241 241 241 241 241 241 241 241 241 241 241 241
241 241 241 241 241 241 241 242 242 242 242 242 242 242 242 242 242 242
242 242 242 242 242 242 242 242 243 243 243 243 243 243 243 243 243 243
243 243 243 243 243 243 243 243 243 243 243 243 244 244 244 244 244 244
244 244 244 244 244 244 244 244 244 244 244 244 244 244 244 244 244 244
244 245 245 245 245 245 245 245 245 246 246 246 246 246 246 246 246 246
246 246 246 246 246 246 246 246 247 247 247 247 247 247 247 247 247 247
247 247 247 247 247 247 247 247 247 247 248 248 248 248 248 248 248 248
248 248 248 248 249 249 249 249 249 249 249 249 249 249 249 250 250 250
250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 251 251
251 251 251 251 251 251 251 251 251 251 251 252 252 252 252 252 252 252
252 252 252 252 252 252 252 252 252 252 252 253 253 253 253 253 253 253
253 253 253 253 253 253 253 253 254 254 254 254 254 254 254 254 254 254
254 254 254 254 254 254 254 254 254 254 254 255 255 255 255 255 255 255
255 255 255 255 255 255 255 255 255 255 255 256 256 256 256 256 256 256
256 256 256 257 257 257 257 257 257 257 257 257 257 257 257 257 257 257
257 257 257 257 257 257 257 257 258 258 258 258 258 258 258 258 258 258
259 259 259 259 259 259 259 259 259 259 259 259 259 259 259 259 260 260
260 260 260 260 260 260 261 261 261 261 261 261 261 261 261 261 261 261
262 262 262 262 262 262 262 262 262 262 262 262 262 262 262 262 263 263
263 263 263 263 263 263 263 263 263 263 263 263 263 263 263 264 264 264
264 264 264 264 264 264 264 264 265 265 265 265 265 265 265 265 265 265
266 266 266 266 266 266 266 266 266 267 267 267 267 267 267 267 268 268
268 268 268 268 268 268 268 268 268 268 268 269 269 269 269 269 269 269
269 269 269 269 269 269 270 270 270 270 270 270 270 270 270 270 270 270
270 270 270 270 270 271 271 271 271 271 271 271 271 271 271 272 272 272 272
272 272 272 272 273 273 273 273 273 273 273 274 274 274 274 274 274 274
274 274 274 274 275 275 275 275 276 276 276 277 277 278 278 278 278 278
278 278 278 278 278 278 279 279 279 279 279 279 279 280 280 280 280 280
280 280 280 280 280 281 281 281 281 281 282 282 282 283 283 283 283 283
283 283 284 284 284 284 284 284 284 284 285 285 285 285 285 286 286 286
286 287 287 287 287 287 287 288 288 288 288 289 290 290 290 290 291 292
292 292 292 293 293 294 294 295 296 296 296 297 297 297 298 298 299 299
299 300 301 301 302 302 304 305 307 308 310 310 313 313 313 313 313 315
315 317 318 319 320 325 326 326 327 328 331 334 334 336 338 339 341 341
342 344 345 345 350 352 354 356 358

References

Aigner, D. J. & Chu, S. F. (1968) On estimating the industry production function. *American Economic Review*, **13**, 568-598.

Aigner, D. J., Lovell, C. A. K & Schmidt, P. (1977) Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, **6**, 21-37.

Aitchison, J. (1964) Confidence-region tests. *Journal of the Royal Statistical Society*, Series B, **26**, 462-476.

Aitchison, J. (1965) Likelihood-ratio and confidence-region tests. *Journal of the Royal Statistical Society*, Series B, **27**, 245-250.

Aitkin, M., Anderson, D., Francis, B. & Hinde, J. (1989) *Statistical Modelling in GLIM*. Oxford: Oxford University Press.

Andel, J., Netuka, Y. & Zvára, K. (1984) On threshold autoregressive processes. *Kybernetika*, **20**, 89-106.

Antille, A., Kersting, G. & Zucchini, W. (1982) Testing symmetry. *Journal of the American Statistical Association*, **77**, 639-646.

Arnold, B. C., Balakrishnan, N. & Nagaraja, H. N. (1992) *A First Course in Order Statistics*. New York: Wiley.

Arnold, B. C. & Beaver, R. J. (2000) Hidden truncation models. *Sankhya*, Series A, **62**, 23-35.

Arnold, B. C. & Beaver, R. J. (2002) Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, **11**, 7-54.

Arnold, B. C., Beaver, R. J., Groeneveld, R. A. & Meeker, W. Q. (1993) The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, **58**, 471-488.

Arnold, B. C. & Shavelle, R. M. (1998) Joint confidence sets for the mean and variance of a normal distribution. *The American Statistician*, **52**, 133-140.

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.

Azzalini, A. (1986) Further results on a class of distributions which includes the normal ones. *Statistica*, **46**, 199-208.

Azzalini, A. & Capitanio, A. (1999) Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society*, Series B, **61**, 579-602.

Azzalini, A. & Dalla Valle, A. (1996) The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

Barnard, G. A. (1963) Discussion of "The spectral analysis of point processes", by M. S. Bartlett. *Journal of the Royal Statistical Society*, Series B, **25**, 294.

Barnett, V. D. (1966) Evaluation of the maximum likelihood estimator where the likelihood equation has multiple roots. *Biometrika*, **53**, 151-165.

Batschelet, E. (1965) *Statistical Methods for the Analysis of Problems in Animal Orientation and Certain Biological Rhythms*. Washington DC: American Institute of Biological Sciences.

Batschelet, E. (1981) *Circular Statistics in Biology*. London: Academic Press.

Bhattacharya, P. K., Gastwirth, J. L. & Wright, A. L. (1982) Two modified Wilcoxon tests for symmetry about an unknown location parameter. *Biometrika*, **69**, 377-382.

Birnbaum, Z. W. (1950) Effect of linear truncation on a multinormal population. *The Annals of Mathematical Statistics*, **21**, 272-279.

Bishop, B. V. (1947) The frequency of thunderstorms at Kew observatory. *Meteorological Magazine*, **76**, 108-111.

Boos, D. D. (1982) A test for asymmetry associated with the Hodges-Lehmann estimator. *Journal of the American Statistical Association*, **77**, 647-651.

Brooks, S. P., Morgan, B. J. T., Ridout, M. S. & Pack, S. E. (1997) Finite mixture models for proportions. *Biometrics*, **53**, 1097-1115.

Bruderer, B. & Jenni, L. (1990) Migration across the Alps. In *Bird Migration: Physiology and Ecophysiology* (ed. E. Gwinner), pp. 60-77. Berlin: Springer-Verlag.

Butler, C. C. (1969) A test for symmetry using the sample distribution function. *Annals of Mathematical Statistics*, **40**, 2209-2210.

Cabilio, P. & Masaro, J. (1996) A simple test of symmetry about an unknown median. *Canadian Journal of Statistics*, **24**, 349-361.

Carnahan, B., Luther, H. A. & Wilks, J. O. (1969) *Applied Numerical Methods*. New York: Wiley.

Catchpole, E. A. & Morgan, B. J. T. (1997) Detecting parameter redundancy. *Biometrika*, **84**, 187-196.

Chatfield, C. (1995) *Problem Solving: a Statistician's Guide* (2nd edition). London: Chapman & Hall.

Cheng, R. C. H. & Traylor, L. (1995) Non-regular maximum likelihood problems. *Journal of the Royal Statistical Society*, Series B, **57**, 3-44.

Chiogna, M. (1997) *Notes on estimation problems with scalar skew-normal distributions*. Technical report 1997.15, Department of Statistical Sciences, University of Padua, Italy.

Cohen, J. P. & Menjoge, S. S. (1988) One-sample runs test of symmetry. *Journal of Statistical Planning and Inference*, **18**, 93-100.

Cook, R. D. & Weisberg, S. (1994) *An Introduction to Regression Graphics*. New York: Wiley.

Cox, D. R. (1975) Discussion of "Statistics of directional data", by K. V. Mardia. *Journal of the Royal Statistical Society*, Series B, **37**, 380-381.

Cox, D. R. (1992) Discussion of "Constrained Monte Carlo maximum likelihood for dependent data", by C. J. Geyer & E. A. Thompson. *Journal of the Royal Statistical Society*, Series B, **54**, 688.

Csörgö, S. & Heathcote, C. R. (1987) Testing for symmetry. *Biometrika*, **74**, 177-184.

D'Agostino, R. B. (1970) Transformations to normality of the null distribution of g_1. *Biometrika*, **57**, 679-681.

Daniel, C. (1959) Use of half normal plots in interpreting factorial two-level experiments. *Technometrics*, **1**, 311-341.

David, H. A. (1981) *Order Statistics*. New York: Wiley.

Davis, C. E. & Quade, D. (1978) *U*-statistics for skewness or symmetry. *Communications in Statistics*, **A7**, 413-418.

Davison, A. C. & Hinkley, D. V. (1997) *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

Dyck, H. D. & Mattice, W. A. (1941) A study of excessive rainfalls. *Monthly Weather Review*, **69**, 293-302.

Efron, B. (1979) Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1-26.

Elandt, R. C. (1961) The folded normal distribution: Two methods of estimating parameters from moments. *Technometrics*, **3**, 551-562.

Ellis, T. M. R. (1990) *Fortran 77 Programming: with an introduction to the Fortran 90 standard* (2nd edition). Wokingham: Addison Wesley.

Finch, S. J. (1977) Robust univariate test of symmetry. *Journal of the American Statistical Association*, **72**, 387-392.

Fisher, N. I. (1993) *Statistical Analysis of Circular Data*. Cambridge: Cambridge University Press.

Fisher, N. I. & Hall, P. (1989) Bootstrap confidence regions for directional data. *Journal of the American Statistical Association*, **84**, 996-1002.

Fisher, N. I. & Hall, P. (1990) New statistical methods for directional data - I. Bootstrap comparison of mean directions and the fold test in palaeomagnetism. *Geophysical Journal International*, **101**, 305-313.

Fisher, N. I., Hall, P., Jing, B-Y. & Wood, A. T. A. (1996) Improved pivotal methods for constructing confidence regions with directional data. *Journal of the American Statistical Association*, **91**, 1062-1070.

Fisher, N. I. & Lewis, T. (1983) Estimating the common mean direction of several circular or spherical distributions with differing dispersions. *Biometrika*, **70**, 333-341.

Fisher, N. I. & Powell, C. McA. (1989) Statistical analysis of two-dimensional palaeocurrent data: methods and examples. *Australian Journal of Earth Sciences*, **36**, 91-107.

Fraser, D. A. S. (1957) Most powerful rank-type tests. *Annals of Mathematical Statistics*, **28**, 1040-1043.

Gastwirth, J. L. (1965) Percentile modification of two sample rank tests. *Journal of the American Statistical Association*, **60**, 1127-1141.

Gastwirth, J. L. (1971) On the sign test for symmetry. *Journal of the American Statistical Association*, **66**, 821-823.

Gibbons, J. F. & Mylroie, S. (1973) Estimation of impurity profiles in ion-implanted amorphous targets using joined half-Gaussian distributions. *Applied Physics Letters*, **22**, 568-569.

Gupta, A. K. & Chen, T. (2001) Goodness-of-fit tests for the skew-normal distribution. *Communications in Statistics - Computing and Simulation*, **30**, 907-930.

Gupta, M. K. (1967) An asymptotically nonparametric test of symmetry. *Annals of Mathematical Statistics*, **38**, 849-866.

Hájek, J. & Sidák, Z. (1967) *Theory of Rank Tests*. New York: Academic Press.

He, X. & Simpson, D. G. (1992) Robust direction estimation. *Annals of Statistics*, **20**, 351-369.

Henze, N. (1986) A probabilistic representation of the 'skew-normal' distribution. *Scandinavian Journal of Statistics*, **13**, 271-275.

Hill, D. L. & Rao, P. V. (1977) Tests of symmetry based on Cramér-von Mises statistics. *Biometrika*, **64**, 489-494.

Hisada, M. (1972) Azimuth orientation of the dragonfly (*Sympetrum*). In *Animal Orientation and Navigation* (eds. S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs and R. E. Belleville), pp. 511-522. Washington DC: U.S. Government Printing Office.

Jammalamadaka, S. R. & Sengupta, A. (2001) *Topics in Circular Statistics*. Singapore: World Scientific.

Jandar, R. (1957) Die optische Rictungsorientierung der roten Waldameise (*Formica rufa* L.). *Zeitschrift für Vergleichende Physiologie*, **40**, 162-238.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994) *Continuous Univariate Distributions,* Vol. 1 (2nd edition). New York: Wiley.

Jones, M. C. (2001) A skew t distribution. In *Probability and Statistical Models With Applications* (eds. C. A. Charalambides, M. V. Koutras & N. Balakrishnan), pp. 269-278. London: Chapman & Hall/CRC.

Jones, M. C. (2002) Marginal replacement in multivariate densities, with application to skewing spherically symmetric distributions. *Journal of Multivariate Analysis*, **81**, 85-99.

Jones, M. C. & Faddy, M. J. (2002) A skew extension of the t distribution, with applications. *Journal of the Royal Statistical Society*, Series B. To appear.

Jupp, P. E. & Spurr, B. D. (1983) Sobolev tests for symmetry of directional data. *Annals of Statistics*, **11**, 1225-1231.

Kappenman, R. F. (1988) Detection of symmetry or lack of it and applications. *Communications in Statistics - Theory and Methods*, **17**, 4163-4177.

Kendall, M. G. & Stuart, A. (1963) *The Advanced Theory of Statistics*, Vol. 1. London: Griffin.

Ko, D. & Guttorp, P. (1988) Robustness of estimators for directional data. *Annals of Statistics*, **16**, 609-618.

Koziol, J. A. (1983) Tests for symmetry about an unknown value based on the empirical distribution function. *Communications in Statistics - Theory and Methods*, **12**, 2823-2846.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. New York: Wiley.

Lenth, R. V. (1981) Robust measures of location for directional data. *Technometrics*, **23**, 77-81.

Lindsey, J. K. (1996) *Parametric Statistical Inference*. Oxford: Oxford University Press.

Liseo, B. (1990) The skew-normal class of densities: inferential aspects from a Bayesian viewpoint (in Italian). *Statistica*, **50**, 59-70.

Loperfido, N. (2002) Statistical implications of selectively reported inferential results. *Statistics and Probability Letters*, **56**, 13-22.

Mardia, K. V. (1971) Discussion of "Spline transformations: three new diagnostic aids for the statistical analyst", by L. I. Boneva, D. G. Kendall & I. Stefanov. *Journal of the Royal Statistical Society*, Series B, **33**, 50-51.

Mardia, K. V. (1972) *Statistics of Directional Data*. London: Academic Press.

Mardia, K. V. & Jupp, P. E. (1999) *Directional Statistics*. Chichester: Wiley.

Matthews, G. V. T. (1961) "Nonsense" orientation in mallard *ans platyrhynchos* and its relation to experiments on bird navigation. *Ibis*, **103**a, 211-230.

McLachlan, G. & Peel, D. (2000) *Finite Mixture Models*. New York: Wiley.

McLeod, A. I. (1985) A remark on AS 183. An efficient and portable pseudo-random number generator. *Applied Statistics*, **34**, 198-200.

McWilliams, T. P. (1990) A distribution-free test for symmetry based on a runs statistic. *Journal of the American Statistical Association*, **85**, 1130-1133.

Merkel, F. W. & Fischer-Klein, K. (1973) Winkelkompensation bei Zwergwachteln (*Excalfactoria chinensis*). *Die Vogelwarte*, **27**, 39-50.

Modarres, R. & Gastwirth, J. L. (1996) A modified runs test for symmetry. *Statistics and Probability Letters*, **31**, 107-112.

Modarres, R. & Gastwirth, J. L. (1998) Hybrid test for the hypothesis of symmetry. *Journal of Applied Statistics*, **25**, 777-783.

Mood, A. M. (1950) *Introduction to the Theory of Statistics*. New York: McGraw-Hill.

Mukhopadhyay, S. & Vidakovic, B. (1995) Efficiency of linear Bayes rules for a normal mean: skewed priors class. *The Statistician*, **44**, 389-397.

Nelder, J. A. & Mead, R. (1965) A simplex method for function minimization. *The Computer Journal*, **7**, 308-313.

Nelson, L. S. (1964) Query 2: The sum of values from a normal and a truncated normal distribution. *Technometrics*, **6**, 469-471.

Neyman, J. (1937) Outline of a theory of statistical estimation based on the classical theory of probability. *Phiosophical Transactions of the Royal Society*, **236**, 333-380.

O'Hagan, A. & Leonard, T. (1976) Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, **63**, 201-203.

O'Neill, R. (1971) Algorithm AS 47: function minimization using a simplex procedure, *Applied Statistics*, **20**, 338-345.

Orlov, A. I. (1972) On testing the symmetry of distributions. *Theory of Probability and its Applications*, **17**, 357-361.

Owen, A. B. (2001) *Empirical Likelihood*. New York: Chapman & Hall/CRC.

Owen, D. B. (1956) Tables for computing bivariate normal probabilities. *Annals of Mathematical Statistics*, **27**, 1075-1090.

Papakonstantinou, V. (1979) Beiträge zur zirkulären Statistik. PhD dissertation. University of Zurich, Switzerland.

Pearson, E. S. (1963) Some problems arising in approximating to probability distributions, using moments. *Biometrika*, **50**, 95-111.

Pearson, K. (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophical Magazine*, **5**, 157-175.

Pewsey, A. (2000a) Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, **27**, 859-870.

Pewsey, A. (2000b) The wrapped skew-normal distribution on the circle. *Communications in Statistics – Theory and Methods*, **29**, 2459-2472.

Pewsey, A. (2002a) Large-sample inference for the general half-normal distribution. *Communications in Statistics – Theory and Methods*, **31**, 1045-1054.

Pewsey, A. (2002b) Testing circular symmetry. *The Canadian Journal of Statistics*. To appear.

Psarakis, S. & Panaretos, J. (1990) The folded t distribution. *Communications in Statistics – Theory and Methods*, **19**, 2717-2734.

Purkayastha, S. (1995) An almost sure representation of sample circular median. *Journal of Statistical Planning and Inference*, **46**, 77-91.

Ramberg, J. S. & Schmeiser, B. W. (1974) An approximate method for generating asymmetric random variables. *Communications of the ACM*, **17**, 78-82.

Randles, R. H., Fligner, M. A., Policello, G. E., & Wolfe, D. A. (1980) An asymptotically distribution-free test for symmetry versus asymmetry. *Journal of the American Statistical Association*, **75**, 168-172.

Roberts, C. (1966) A correlation model useful in the study of twins. *Journal of the American Statistical Society*, **61**, 1184-1190.

Rothman, E. D. & Woodroofe, M. (1972) A Cramér-von Mises type statistic for testing symmetry. *Annals of Mathematical Statistics*, **43**, 2035-2038.

Salvan, A. (1986) Locally most powerful invariant tests of normality (in Italian). In *Atti della XXXIII Riunione Scientifica della Società Italiana di Statistica*, **2**, pp. 173-9. Bari: Cacucci.

Schach, S. (1969) Nonparametric symmetry tests for circular distributions. *Biometrika*, **56**, 571-577.

Schmidt, P. (1976) On the statistical estimation of parametric frontier production functions. *Review of Economics & Statistics*, **58**, 238-239.

Schmidt, W. (1917) Statistische Methoden beim Gefügestudium Kristalliner Schiefer, *Sitz. Kaiserl. Akad. Wiss. Wien, Math. – nat. Kl. Abt.* 1, **126**, 515-538.

Schuster, E. F. & Barker, R. C. (1987) Using the bootstrap in testing symmetry versus asymmetry. *Communications in Statistics – Simulation and Computation*, **16**, 69-84.

Sengupta, A. & Pal, C. (2001) On optimal tests for isotropy against the symmetric wrapped stable – circular uniform mixture family. *Journal of Applied Statistics*, **28**, 129-143.

Sengupta, S. & Rao, J. S. (1966) Statistical analysis of cross-bedding azimuths from the Kamthi formation around Bheemaram, Pranhita-Godavari valley. *Sankhyā* B, **28**, 165-174.

Smirnov, N. V. (1947) On criteria for the symmetry of distribution laws of random variables. *Rossiiskaya Akademiya Nauk*, **56**, 13-16.

Smith, R. L. (1985) Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, **72**, 67-92.

Smith, R. L. (1989) A survey of nonregular problems. In *Proceedings of The International Statistical Institute Conference, 47th Session, Paris,* pp. 353-372.

Smith, R. L. & Naylor, J. C. (1987) A comparison of maximum likelihood and bayesian estimators for the three-parameter Weibull distribution. *Applied Statistics,* **36**, 358-369.

Spanier, J. & Oldham, K. B. (1987) *An Atlas of Functions*. Washington DC: Hemisphere.

Srinivasan, R. & Godio, L. B. (1974) A Cramér-von Mises type statistic for testing symmetry. *Biometrika*, **61**, 196-198.

Stephens, M. A. (1969) *Techniques for directional data*. Technical report 150, Department of Statistics, Stanford University.

Tajuddin, I. H. (1994) Distribution-free test for symmetry based on Wilcoxon two-sample test. *Journal of Applied Statistics*, **21**, 409-415.

Weinstein, M. A. (1964) Query 2: The sum of values from a normal and a truncated normal distribution. *Technometrics*, **6**, 104-105.

Wichmann, B. A. & Hill, I. D. (1982) Algorithm AS 183: an efficient and portable pseudo-random number generator. *Applied Statistics*, **31**, 188-190.

Wilmore, J. H. & Costill, D. L. (1994) *Physiology of Sport and Exercise.* Champaign, Illinois: Human Kinetics.

Winter, A. (1947) On the shape of the angular case of Cauchy's distribution curves. *Annals of Mathematical Statistics*, **18**, 589-593.

Yeh, A. B. & Singh, K. (1997) Balanced confidence regions based on Tukey's depth and the bootstrap. *Journal of the Royal Statistical Society*, Series B, **59**, 639-652.

Yule, G. U. (1911) *Introduction to the Theory of Statistics*. London: Griffin.